Subject SP8

CMP Upgrade 2022/23

CMP Upgrade

This CMP Upgrade lists the changes to the Syllabus, Core Reading and the ActEd material since last year that might realistically affect your chance of success in the exam. It is produced so that you can manually amend your 2022 CMP to make it suitable for study for the 2023 exams. It includes replacement pages and additional pages where appropriate.

Alternatively, you can buy a full set of up-to-date Course Notes / CMP at a significantly reduced price if you have previously bought the full-price Course Notes / CMP in this subject. Please see our 2023 *Student Brochure* for more details.

We only accept the current version of assignments for marking, *ie* those published for the sessions leading to the 2023 exams. If you wish to submit your script for marking but only have an old version, then you can order the current assignments free of charge if you have purchased the same assignments in the same subject in a previous year, and have purchased marking for the 2023 session.

This CMP Upgrade contains:

- all significant changes to the Syllabus and Core Reading
- additional changes to the ActEd Course Notes and Assignments that will make them suitable for study for the 2023 exams.

1 Changes to the Syllabus

This section contains all the ***non-trivial*** changes to the syllabus objectives.

Objectives 3.8 has been reworded as follows, to include reference to machine learning:

3.8 Understand generalised linear models, multivariate modelling and machine learning techniques.

3.8.1 Assess the applications of generalised linear models to the rating of personal lines business and small commercial risks.

3.8.2 Outline the different types of multivariate models and machine learning techniques.

3.8.3 Evaluate the uses of multivariate models and machine learning techniques in pricing.

2 Changes to the Core Reading and ActEd text

This section contains all the ***non-trivial*** changes to the Core Reading and ActEd text.

Chapter 1

Section 8.1

The first paragraph at the top of page 26 has been reworded to aid clarity as follows:

> This is the amount of claims paid (as above) *plus* the increase in the total reserve for outstanding claims (also expressed as claims incurred). For example, suppose that a claim for £1,000 is reported on 20 December 2018, but payment is delayed until 2 January 2019. In this case, the 2018 claims *paid* would be unchanged but the 2018 reserve for outstanding claims would increase by £1,000, leading to an increase of £1,000 in the 2018 claims *incurred*.

Chapter 6

Section 1.1

The first sentence in the last paragraph at the bottom of page 3 has been deleted, in line with the change to the Glossary definition of 'return commission' outlined below.

Chapter 7

Section 3

Additional material has been added to this section. Please use replacement pages 19 to 22a provided at the end of this Upgrade to update your materials accordingly.

Chapter 8

Section 1.2

On page 4, immediately after the first paragraph of Core Reading, the following additional paragraph of Core Reading has now been added:

> **It is also important to consider the time period over which inflation will be applied. Within a pricing context, although we will typically price contracts for the forthcoming year, the inflation assumptions need to reflect the claims environment at the point of settlement of the claims arising from the policy period. This may be many years after the initial policy was written, so could experience much heavier inflation than at the point of writing the policy. Similarly, when reserving, the inflation assumptions need to reflect expectations over the period to settlement of the claims, which will depend on the maturity of the portfolio.**

Chapter 13

Section 2

Significant changes have been made to Section 2.3, and some changes have also been made to Section 2.6. Please use replacement pages 11 to 16 provided at the end of this Upgrade to update your materials accordingly.

Summary

The summary at the end of the chapter has been updated in light of the above changes.

Chapter 16

Chapter 16 has been renamed as 'Generalised linear modelling and machine learning'. Two new sections have been added to this chapter, which are now Section 1 and Section 7 (and the other sections have been renumbered accordingly). These new sections discuss machine learning in a wider context, beyond simply GLMs. Please use replacement pages 1 to 2c and replacement pages 54a to 54b provided at the end of this Upgrade to update your materials accordingly.

Summary

The summary at the end of the chapter has been updated in light of the above changes.

Practice Questions

The following two new practice questions have been added:

16.8 State the two main classes of modelling problems within supervised machine learning.

16.9 State whether each of the following applications involves supervised or unsupervised learning:

(a) predicting a motor insurance policyholder's expected claim frequency based on their age, postcode, car make and the number of years for which they have held their licence

(b) identifying groups of insurance products that tend to be bought together.

Solutions

The corresponding solutions have been added as follows:

16.8 The two main classes of modelling problems within supervised machine learning are:

- classification problems – these involve associating a given input with a particular label or class, *eg* distinguishing apples from oranges
- regression problems – these involve predicting a continuous variable such as expected frequency or severity of claims.

16.9 (a) This would be supervised learning – we would aim to reproduce the observed claim frequencies for a sample of policyholders as closely as possible based on the four rating factors, *eg* using a GLM.

(b) This would be unsupervised learning – we hope that the algorithm will identify clusters of products that tend to be bought together, but these are unknown in advance.

Glossary

The definition of 'return commission' has been amended, to remove 'Also called overriding commission.' The revised definition is therefore simply:

Commission paid by a reinsurer to an insurer ceding proportional business, as a contribution towards expenses and profit.

3 Changes to the X Assignments

Overall

There have been a number of minor improvements to the X Assignment questions and solutions. Due to the number of changes involved, we have not detailed them individually in this upgrade, and recommend that the new 2023 X Assignments should be used.

If you would like the new assignments *without* marking, then retakers can purchase an updated CMP or standalone X Assignments at a significantly reduced price. Further information on retaker discounts can be found at:

acted.co.uk/paper_reduced_prices.html

If you wish to submit your scripts for marking but only have an old version, then you can order the current assignments free of charge if you have purchased the same assignments in the same subject in a previous year, and have purchased marking for the 2023 session. We only accept the current version of assignments for marking, *ie* those published for the sessions leading to the 2023 exams.

4 Other tuition services

In addition to the CMP you might find the following services helpful with your study.

4.1 Study material

We also offer the following study material in Subject SP8:

- Flashcards
- Revision Notes
- ASET (ActEd Solutions with Exam Technique) and Mini-ASET
- Mock Exam and AMP (Additional Mock Pack).

For further details on ActEd's study materials, please refer to the 2023 *Student Brochure*, which is available from the ActEd website at **ActEd.co.uk**.

4.2 Tutorials

We offer the following (face-to-face and/or online) tutorials in Subject SP8:

- a set of Regular Tutorials (lasting a total of three days)
- a Block (or Split Block) Tutorial (lasting three full days)
- an Online Classroom.

For further details on ActEd's tutorials, please refer to our latest *Tuition Bulletin*, which is available from the ActEd website at **ActEd.co.uk**.

4.3 Marking

You can have your attempts at any of our assignments or mock exams marked by ActEd. When marking your scripts, we aim to provide specific advice to improve your chances of success in the exam and to return your scripts as quickly as possible.

For further details on ActEd's marking services, please refer to the 2023 *Student Brochure*, which is available from the ActEd website at **ActEd.co.uk**.

4.4 Feedback on the study material

ActEd is always pleased to receive feedback from students about any aspect of our study programmes. Please let us know if you have any specific comments (*eg* about certain sections of the notes or particular questions) or general suggestions about how we can improve the study material. We will incorporate as many of your suggestions as we can when we update the course material each year.

If you have any comments on this course, please send them by email to **SP8@bpp.com**.

3 Fiscal regimes

3.1 The need for supervision

Why should insurance business suffer more legislation than, say, umbrella manufacturers? One of the reasons is that there is more scope for the purchaser to lose out financially. When you buy an umbrella, you have a look at it, and if you like it, you pay the price. However, with insurance, you pay the price at the start of the contract and you have to trust the insurer to pay valid claims as and when they arise in the future.

The uncertainty underlying insurance business means that it is not just a question of trusting the honesty of the insurer. The insurer may be very well meaning, but if the insurer's business is not soundly managed, you may find that the insurer has collapsed by the time you need to make a claim.

In many countries, therefore, there are specific rules and regulations that apply to general insurers. Different countries adopt different approaches to the regulation of insurers' operations.

3.2 Effect of the regulatory regime

The following regulatory restrictions on the actions of a general insurer may be encountered in one or more countries of the world:

- **A requirement to have an external audit of the general insurer's accounts and to release publicly certain information pertaining to the insurer.**
- **Restrictions on the type of business that a general insurer can write or classes for which the insurer is authorised.** An authority could prevent an insurer from writing volatile classes of business or classes where it has little expertise.
- **Limits or controls on the premium rates that can be charged.**

  For example, the authorities in some US states, *eg* Massachusetts, set the personal motor premium rates that must be charged. Some states require that rates are filed (sent to the relevant state department for approval) prior to an insurer using them. An authority could also set a maximum or minimum premium or restrict the way in which the premiums are calculated. For example an authority could set a maximum allowance for expenses defined as a percentage of the gross premium.
- **Restrictions on the information that may be used in underwriting and premium rating,** perhaps to avoid unfair discrimination. For example, under the EU Gender Directive, European insurers cannot use gender as a rating factor. (This is discussed further below.)
- **A requirement to ensure fairness in the treatment of customers.**
- **A requirement to deposit assets to back claims reserves.**

- **A requirement that the general insurer maintains a minimum level of solvency, measured in some prescribed manner,** *ie* a minimum level of free assets. This might, for example, be calculated as a proportion of premiums written.
- **Restrictions on the types of assets or the amount of a particular asset that a general insurer can take into account for the purposes of demonstrating solvency.** This might be with the possible aim of avoiding risky investments or increasing diversification.
- **A requirement to use prescribed bases for calculating premiums or for valuing the general insurer's assets and/or liabilities when demonstrating solvency.**
- **Restrictions on individuals holding key roles in companies.**
- **Licensing of agents to sell insurance and requirements on the methods of sale and disclosure of commission / broking terms.**
- **A requirement to respond to thematic reviews conducted by the regulator.**
- **A requirement to pay levies to consumer protection bodies.**
- **Legislation to protect policyholders if a general insurer fails.**

Question

Suggest possible legislation that could be used to protect policyholders if a general insurer fails.

Solution

A fund could be set up to pay claims from failed general insurers. It could be funded by the government or by charging a levy on other insurers.

Companies could be required to deposit a large initial sum to a central governing body. This could be used to pay claims on default.

Fairness within the context of insurance

Regulators are keen to ensure that insurance products provide value to the customers and that customers are treated fairly. There are various ways that this can be addressed, from guidance on acceptable practices and rating approaches (see examples below), through rate filings and even to provision of cover itself, where the commercial markets would make insurance purchase prohibitively expensive (*eg* buying household flood insurance for a property that is built on a flood plain).

An example of a regulatory restriction: EU Gender Directive

The EU Gender Directive was passed in 2004, being aimed at 'implementing the principle of equal treatment between men and women in the access to and supply of goods and services'.

In its original form, the EU Gender Directive included an opt-out in respect of financial and insurance products provided that certain conditions were met. In March 2011, the European Court of Justice gave its ruling on the legality of the insurance opt-out provision, concluding that it is not valid and should therefore be removed with effect from 21 December 2012. From that point, insurance companies have no longer been able to use gender as a rating factor.

Insurance companies are careful to avoid the use of proxy rating factors (*ie* highly correlated to gender) that might be deemed to be indirect discrimination and thus also not permitted.

Clearly, the inability to differentiate between gender when setting premium rates is having significant implications for insurance pricing, particularly for motor insurance where there are material observed differences in claims experience according to gender at certain ages.

Each insurer is likely to set premium rates based on the expected mix of business by gender but there is the risk that this mix turns out not to be as expected. The introduction of this legislation has therefore increased the uncertainty of insurers' claims experience and profitability.

It is not yet clear how premium rates or underwriting practices have changed as a result of the ruling. However it is likely that premiums have not simply 'met in the middle', but that there have been additional contingency loadings for the risk of business mix by gender not being as expected within the unisex pricing.

In other words, this legislation has also led to increased uncertainty, at least in the short term, and hence higher risk margins being incorporated in premiums by insurers.

An example of regulator thematic reviews: FCA Review of UK general insurance pricing practices

During 2019 and 2020, the Financial Conduct Authority (FCA), a UK regulator, conducted a review of the UK home and motor insurance markets. The FCA concluded that these markets were not working well for consumers. In particular, loyal customers were not getting good value and the FCA estimated that six million policyholders paid prices that were too high in 2018 and that if they had paid the average for their risk, they would have saved £1.2*bn*.

The FCA will be prohibiting any differential in pricing between new business and renewal policies going forward and banning the use of 'price walking' where customers pay a lower premium in their first year and their prices increase over time.

This legislation came into force in the UK from 1 January 2022. UK home and motor insurers now have to charge a renewing policyholder the same premium as they would charge an equivalent new policyholder.

The FCA also want to make it easier for customers to switch insurers. Policies that automatically renew discourage customers from shopping around. The FCA are looking for the insurance industry to meet customers' needs and in doing so treat them fairly. Technology is seen as a barrier to some customers and, to date, has meant it is easier for some customers to shop around to achieve lower prices.

In future, the FCA wants to provide long-term fair value for all customers and for customers to be able to trust this. There will be more monitoring by the FCA to ensure the new rules are being followed.

Other possible regulations

Other regulations that could be imposed on general insurers include:

- requirement to provide detailed reports and accounts at prescribed intervals
- requirement to purchase reinsurance
- requirement to hold a claims equalisation reserve
- limits on contract terms
- advertising restrictions
- prescription to hold certain assets.

Regulation is considered further in Subject SP7.

3.3 Effect of the fiscal regime

In most countries, the taxation of general insurers broadly follows that for other businesses although there may be special features, such as allowing equalisation reserves to be held to allow for the uncertain nature of general insurance business.

For example, in some countries, transfers to equalisation reserves or catastrophe reserves may be allowable against taxable profit.

Some countries impose a tax on general insurance premiums for some or all classes of business.

4 Professional guidance

When carrying out work for a general insurer or reinsurer an actuary should always bear in mind any professional guidance relating to the work being carried out and the professional body to which they belong.

Institute and Faculty of Actuaries (IFoA) members practising in the UK need to comply with the Technical Actuarial Standards (TASs) that are issued by the Financial Reporting Council. Although the content of the TASs is not officially part of the Core Reading, a familiarity with them may help you pass the exam.

An actuary should also bear in mind guidance on professional standards in addition to guidance on technical issues.

All members of the IFoA are subject to The Actuaries' Code. This gives general guidance on professional conduct to which all IFoA members must conform in both the spirit and the letter.

In addition to formal guidance, the professional body may issue advice from time to time on specific issues.

The details of all the latest professional guidance issued by the IFoA can be obtained via its website: www.actuaries.org.uk.

Professional guidance is covered in more detail in Subject SA3.

As discussed above, the insurer may charge renewing customers a higher premium than that charged to new customers. If this is the case, it will make more profit from each renewal, compared to new policies. On the other hand, if premiums are getting cheaper due to increased competition, renewing customers may be more tempted to switch insurers. The insurer may decide to reduce premiums, particularly on renewal, in order to encourage customer loyalty.

Conversely, in a hard market, insurers may charge a higher rate than is required.

They can make more profit because there is less competition.

The importance of the insurance cycle in pricing is discussed more fully below, in Section 2.5.

Market acceptability

As mentioned above, the theoretical premium for renewal business may be lower than that for new business.

This was discussed in the previous chapter in the section on expense loadings.

For some classes of business, for example motor insurance, the claims experience for new business can also be significantly worse than for renewal business.

The theoretical premium should therefore be higher for new business than for renewals.

However, charging these different premiums in practice is unlikely to be acceptable to consumers. As a result, the insurer usually charges the same premium for both new business and renewals, with the total costs spread over all contracts.

Use of no-claims discounts

A no-claim discount (NCD) system is where an individual policyholder may be granted a discount from the relevant base premium depending on their claims experience. They are common on private motor insurance, but can be applied to other personal lines products, such as household insurance.

These are often set by market practice rather than being statistically justifiable.

Distribution channels

Different distribution channels may require different strategies for the insurance company.

Although the technical premium is unlikely to be directly affected by the distribution channel, there may be a difference as a result of acquisition costs (*eg* using a broker is likely to increase acquisition costs). There may also be system constraints between distribution channels (*eg* a broker portal may only have a subset of rating factors that an insurer may want to use).

Differences in acquisition costs may affect the technical premium. System constraints on the number of rating factors that can be used may affect the calculated risk premium. Hence both of these factors could affect the actual price charged.

There could also be some underlying proxy to risk that manifests itself through the distribution channel (*eg* an affinity scheme that sells motor insurance to classic car enthusiasts).

In other words, the 'riskiness' of policyholders may differ by distribution channel for some reason. The Core Reading gives the example of sales to an affinity group as a possible reason. An affinity group is simply a group of individuals with something in common. For example, everybody that shops at Sainsbury's would be an affinity group, as would everyone that belongs to a classic car enthusiasts' club, as in the example above.

Effect of legacy business:

An established insurer may need to allow for prior years' reserve deficiencies / surpluses – that is, the need to increase / release reserves for business already on the books. A new entrant to the market will not have such reserves. On the other hand, a new entrant will normally have relatively high expenses until its business matures, but may also have lower overheads to be spread per policy.

New entrants will normally have relatively high expenses because the expenses associated with writing new business are generally much higher than those for renewing business. This means that, in most cases, established insurers will incur lower expenses than new insurers because a lower proportion of their business will be for new (as opposed to renewing) policies. Consequently, established insurers may charge lower premiums. Whilst this is generally true, there are reasons why new insurers may not necessarily charge more than their more established competitors.

Question

A student told a friend: 'There won't ever be any new motor insurers. They wouldn't attract any business because the rates they would have to charge are higher than those charged by an established insurer.'

Comment on this statement.

Solution

The statement is almost certainly not true because:

- a new company may be able to subsidise its premium rates until it has a mature portfolio
- the new company may accept a lower profit margin, especially if it entered the market at the top of the insurance cycle
- the brand name of the company may be attractive to potential customers
- the company may introduce innovative product designs, *eg* new scales of NCD
- a new company may employ good managers and use efficient new processes that result in lower expenses than the established companies
- innovative product design or services may attract customers
- a new company will not need to have its profits depressed by the need to fund any prior year deficits which some established insurers might have.

2.4 Profit optimisation

Insurers can optimise the total profit by accepting a lower profit margin on individual policies in return for a higher total business volume. Although a lower profit is achieved by each policy, the business volume generated by the lower level of premium may result in a higher total profit to the company.

In other words, the profit per policy will be lower, but the total profit across all policies may be higher as a result of selling a greater number of policies. Hence, price elasticity is a key consideration when considering a change in premium rates.

This needs to be balanced against the potentially increased risk of a larger, lower margin portfolio.

We're not talking about price elasticity here. Even if the company succeeds in increasing profit by reducing prices, the resulting portfolio will be more risky (*ie* the profit margin only needs to drop by a small amount to result in a loss).

2.5 Pricing and the insurance cycle

The insurance cycle is a phenomenon that affects all forms of insurance and reinsurance. It is driven by the changing levels of profitability in the market arising from changes in market capacity.

Stages of the insurance cycle

When premium rates are more profitable (hard), new entrants are attracted to the market. This increased capacity causes rates to fall until they may even become loss making. At this point some market participants will leave the market, or reduce their market share. This effect may be exacerbated by catastrophes and economic factors. This reduction in market capacity leads to rates increasing once more until we reach our starting point of the cycle and it begins again.

This was also discussed in Chapter 8.

Reasons for insurance cycles

The cyclical nature of premium rate movements is a very real market challenge, perhaps the most challenging aspect of managing a portfolio of risks for many lines of business.

It is important that the insurer is aware of the stage of the insurance cycle that the market is in when deciding on the final premium.

There are many causes of underwriting cycles, including:

- **a delay in the understanding of the emergence of higher claims costs, which leads to under-pricing of current risks**
- **inflows and outflows of capacity to the market, often as a result of large-scale catastrophes**

  Catastrophes may jolt prices higher due to providers leaving the market hence causing a reduction in supply.

- **deliberate under-pricing (or the use of expense advantage) by key players in an attempt to drive out competitors**
- **an attempt to grow in volume in order to cover high fixed expenses** (or to achieve market share)
- **pricing strategy being determined by chasing market prices (upwards or downwards) rather than being based on sound technical prices, with no player in a given market willing to be the first to break the cycle**
- **the level of investment return available to offset underwriting losses.**

Different markets are impacted to a greater or lesser extent by these factors, and certain markets tend to exhibit more or deeper cyclical trends than others.

Importance of knowing the current stage of the insurance cycle

We should know the stage of the insurance cycle that the market has reached in order to assess how to allow for business objectives and competition. While we should evaluate every risk on its available information (or that pertaining to its class), the tide and optimism of the market can often affect judgement; and indeed, the data itself may be less reliable at certain stages of the market.

Example

In a rapidly rising market, the insurer might give priority to the new business area in the allocation of systems and administration time, with the result that claims information might be less up-to-date.

In addition, we might be less prudent in our claims estimation and reserving, if we believe that the business currently being written is more profitable than it really is.

Dealing with the insurance cycle

Market cycles are a feature of the insurance market, but arguably we can reduce the associated risk, if we base prices on a model with a sound technical cost base and communicate this clearly to the pricing decision-makers. Everyone then clearly recognises and understands departures from the technical price. Actuaries can have a key role to play in both of these areas.

Question

'There is no point in offering an actuarial theoretical premium rate if the rest of the market is undercutting you'. Discuss whether you agree with this comment.

Solution

There is certainly no point in offering higher premium rates than everybody else, if the higher price means that you attract no business and the business would have been profitable at the lower rate.

However, if the business would be unprofitable at the lower rate, then it may be best not to take on the business (although a view will have to be taken on the *long-term* impact of turning down business – for example, it may in the long run be more expensive not to sell business and suffer the later cost of trying to rebuild market share).

There are also differentiators other than price. A more attractive product may still win business despite being more expensive if the added benefits outweigh the extra cost to the policyholder.

Further practical considerations in rating are discussed in the next chapter, when we consider the frequency-severity and burning cost approaches.

There can be a correlation between investment markets and insurance cycles. The circumstances found in many territories in the world during the period following the economic downturn in 2007/8, meant that corrective cycle strategies were often difficult to implement. Features of this period were typically: general financial pessimism, very low financial return available from almost all asset classes, and the need to invest in non-correlated implements such as hedge funds.

2.6 Other influences on rating

In practice there are many other factors that affect the premium rate charged. These include:

- The availability of capital to support new business.

  A large capital base means that riskier products or larger volumes of business can be written. The downside is that, in return for offering greater capital, the providers of that capital will probably require a greater return, and so this may mean higher loadings for profit are needed in the premium rates.

- The impact of reinsurance capacity.

  If reinsurance is not readily available at an acceptable price, the original insurer may have to increase its premium rates to compensate for the extra risk it retains.

- The sophistication of sales and quotes systems.

  Complicated flexible sales and quotes systems cost money, which probably has to be recouped from premium rates eventually.

- The demands of regulators in the rating area.

  In some countries, there are regulatory restrictions on premium rates. This could simply be a requirement to file rates with the regulators (as in parts of the US), or could be restrictions on rates (maxima or minima) or rating factors used (as in some parts of Africa).

- Relationships with particular distributors or brokers.

  For example, in order to appease a particular broker that brings in large volumes of profitable business, a company might wish to tweak prices so that they favour that broker's target market.

16

Generalised linear modelling and machine learning

Syllabus objectives

3.8 Understand generalised linear models, multivariate modelling and machine learning techniques.

- 3.8.1 Assess the applications of generalised linear models to the rating of personal lines business and small commercial risks.
- 3.8.2 Outline the different types of multivariate models and machine learning techniques.
- 3.8.3 Evaluate the uses of multivariate models and machine learning techniques in pricing.

0 Introduction

With the increasing power of computers, the use of generalised linear models (GLMs) is widespread for personal lines pricing; it is used, to a lesser extent, for pricing some commercial lines too. Machine learning techniques more generally are also increasingly being used. There are a number of commercially available software packages that enable actuaries and underwriters to calculate frequencies, average claim costs and burning costs to use as a basis for setting future premium rates.

This chapter focuses primarily on GLMs. It considers some of the mathematics behind GLMs, which you may have met in your earlier studies, and then goes on to consider how we might use these models. If it has been a while since you studied GLMs, it is probably worth reviewing your previous study material before tackling this chapter.

Machine learning is covered in detail in Subject CS2. The use of GLMs for pricing is one example of machine learning and, as mentioned above, is the primary focus of the chapter. However, the chapter also revisits the basics of machine learning more generally and discusses, at a high level, two other machine learning models used by insurers for pricing.

Section 1 provides an overview of machine learning techniques.

Section 2 explains the concepts of GLMs and looks at the exponential family of distributions in some detail.

In Section 3 we describe the principles of how a GLM is constructed and look at the types of factors that might be included within a model.

Section 4 deals with the techniques we can use to check the significance of the factors used in the model.

Section 5 describes techniques that can be used to check the appropriateness of the chosen model structure, *eg* by considering the residuals.

Section 6 considers how we might refine our model. We look at how we can use interactions and offsets, and discuss aliasing of potential rating factors.

Finally in Section 7, we consider two further machine learning models, beyond GLMs, namely gradient boosting models and artificial neural networks.

1 Machine learning – overview

Machine learning refers to a collection of methods for the automatic detection and exploitation of patterns in data. The field has emerged in response to significant increases in the volume of available data, and in the speed and capacity of computers to analyse it.

One goal of any machine learning model is for it to be good or 'performant', where performant is a measure of how well the model can generalise to produce the desired output on previously unseen data. A model will have one or more cost functions that quantitatively measure the algorithm's ability to generalise.

In machine learning, the 'machine' tries to identify the best model within a class of possible models. We tell the machine what a 'good' model looks like by defining a cost function, such that the cost function takes a low value when a model is 'good' and a high value when a model is 'bad'. The aim of the exercise is then to find the model that minimises the cost function.

For example, mean squared error is a cost function for linear regression.

The mean square error (MSE) compares a model's predictions $\hat{f}(x_i)$ with the true output values y_i. It is defined as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{f}(x_i))^2.$$

It is also important to be able to explain the model to management.

There are two main approaches for a model to learn which parameters will produce a useful output:

- **Supervised learning – The model is presented with 'labelled training examples' and a dataset of 'feature' data. The algorithm searches the feature data for patterns that are predictive of the target labels. The model can then use these learned patterns to make predictions on unlabelled data.**

  An example of supervised learning would be the use of generalised linear models for insurance pricing, which will be discussed later in this chapter.

- **Unsupervised learning – An unsupervised model does not require any labelled examples. Instead, it seeks to find latent structure in the data.**

  In other words, there is no output variable to aim at. Instead, the objective is to identify patterns in the input data.

  An example of unsupervised learning is market basket analysis, which identifies items commonly bought together. This can be used by online retailers as the basis for 'Other customers who bought X also bought Y' recommendations or for promoting bundles of items that are frequently bought together.

Most applications within general insurance currently use supervised learning models. Within supervised learning, there are two main classes of modelling problems, namely 'classification' and 'regression':

- **Classification problems involve associating a given input with a particular label or class, for example distinguishing apples from oranges, cats from dogs or fraud from non-fraud. The output can be a class label or a probability.**

  Designing email filters that classify incoming mail as either 'Spam' or 'Not spam' would be an example of a classification problem in everyday life. Not all classification problems are binary in nature, but the most common ones are.

- **Regression problems involve predicting a continuous variable such as expected frequency or severity of claims.**

Notice that, with regression problems, the output variable is always quantitative / numerical whereas, with classification problems, the output variable can be qualitative / categorical.

There are many types and implementations of machine learning algorithms, so expertise and domain knowledge have to be used to decide which is likely to be a good choice for a particular problem.

Domain knowledge refers to specialised knowledge in a particular area, as opposed to general knowledge.

The choice of algorithm is also a trade-off between performance, complexity and the ability to explain or interpret results.

We will generally want the algorithm to be as accurate as possible, without it becoming so complex that it becomes hard to understand or explain the results.

Once the model is chosen, the next step is to find the right model parameters to produce the desired output. This is a complex optimisation task, and typically models will 'learn' via an iterative optimisation procedure.

In other words, the process used to determine the correct parameter values for the model is iterative.

Complex models are not necessarily better models. Although it is important for the output to be a good fit for the training data, it is more important how well the model can generalise on unseen data. This leads to the 'bias variance trade-off':

- **A model with high bias is one which ignores what the data is telling it.**
- **A high-variance model is one where a small change in the training data results in a large change in the model parameters.**

High variance is a sign that the model may have been overfitted to the training data, whereas high bias is a sign that the model is not flexible enough to capture all of the important features of the data.

For example, introducing a new item into the training data may have little effect on a high-bias model, but could have a large effect on a high-variance model. There is usually a trade-off between bias-related error and variance-related error.

Bias and variance are considered 'reducible' errors because they can be reduced by improved modelling. Errors relating to natural variability and the stochastic nature of the system are considered 'irreducible' because they cannot be 'modelled' away.

Irreducible error is essentially another term for what is often called 'noise' in the data. If we were modelling claim severity, for example, this would relate to the inherent randomness in the observed claim sizes. This is unaffected by how good our model is.

However, improved modelling may allow us to better capture the important features of the data and distinguish these from irregularities in the data. This can reduce variance and bias.

In this chapter, we will concentrate on generalised linear models (GLMs), which are a common supervised machine learning technique.

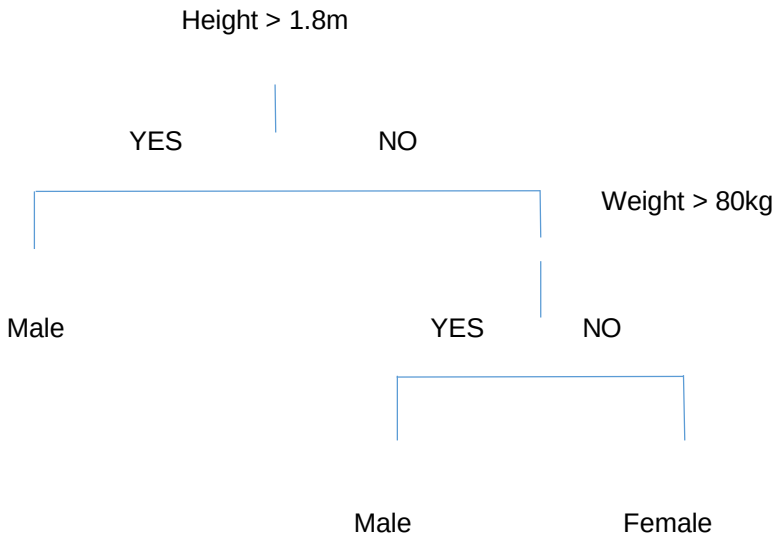
7 Other machine learning models

Whilst GLMs are a powerful technique for an insurer, there are some downsides: in terms of raw predictive power, GLMs are typically less performant out of the box. Some other models are outlined below.

7.1 Gradient boosted models (GBM)

GBMs are decision tree-based models but share many mathematical and conceptual similarities with GLMs and are a natural extension of GLMs. GBMs are part of a class of algorithms called 'ensembles'. An ensemble algorithm combines multiple models together to generate a single algorithm that is more performant than the individual algorithms. In GBMs, each individual algorithm is a decision tree.

A decision tree is an algorithm that uses a series of 'questions' to partition data and arrive at a solution. Visually, it is a branching flowchart that looks like a tree that has been turned upside down. For example:



The example above is a classification tree for predicting gender from the inputs 'height' and 'weight'. (Note that for the purposes of this example, we consider the two genders of male and female, and no other gender classifications.) To predict a person's gender, we begin at the top of the flowchart and, at each branching point, select the appropriate path. Should their height exceed 1.8m, we predict that they are male. If not, we look at their weight: if it exceeds 80kg, we predict that they are male; otherwise, we predict that they are female.

Decision trees were covered in Subject CS2.

The individual performance of any one tree can be minimal, with each tree referred to as a weak learner, but when combined together correctly the overall ensemble becomes a 'strong' learner.

GBMs could have hundreds or thousands of decision trees in the ensemble, and this makes interpreting and explaining the model virtually impossible. There are two separate concepts here:

- **interpretability – the extent to which cause and effect can be observed within the model**
- **explainability – the extent to which the internal mechanics of a model can be explained in human terms.**

Because of these issues, it may be preferable to use a simpler less performant model.

7.2 Deep learning and artificial neural networks

An artificial neural network has a mathematical structure which is loosely inspired by the brain's neurons.

In the brain, neurons are nerve cells that communicate with other cells via specialised connections called synapses.

In an artificial neural network, each neuron takes inputs and applies a linear transformation, of the form $WX + B$, where each input, X, is multiplied by a weight, W, the resulting products are summed, and a bias term B is added. The weights and bias terms are all parameters that are learnt during the model fitting process.

The output of this linear transformation then passes through a non-linear transformation, called an activation function, and the outputs of this activation function are the outputs of the nodes.

A node is another name for a neuron in an artificial neural network. The output of each neuron is found by taking a weighted sum of all its inputs, adding a bias term, and then applying the relevant activation function.

If further layers are added to the network, the neural network is termed 'deep learning'.

The main benefits of artificial neural networks are that:

- **they can be used to tackle a vast array of different problems and offer potentially the best performance of any class of algorithm where there is sufficient data**
- **deep learning models have a reduced need for feature engineering.**

Feature engineering was covered in Subject CS2. It refers to the act of modifying raw data to produce features that can be made use of in supervised learning.

The main disadvantages of these models are that:

- **they require a large amount of data to train the model**
- **deep learning models can be computationally expensive to train**
- **they are truly 'black box' in terms of interpretability.**

  It may be hard to know how the neural network produced a particular output value. For example, the output in question might be the premium rate to charge for a particular segment of customers. This could make it difficult to justify the rate to management or policyholders.