Maths background for MLPR

This course assumes that you have some university-level mathematics experience. For example, as covered by first-year undergraduate mathematics courses taken by informatics, physics, or engineering students in Edinburgh.

Please take the self-assessment test as a guide. You don't have to be able to do all of it. However, the answers should make perfect sense. Moreover, you will need to be able to do future exercises without looking at the answers. If you don't have experience solving technical problems involving mathematics, you are likely to struggle.

It is expected that you are used to manipulating algebraic expressions, and solving for unknowns. For example, it should be straightforward for you to rearrange an expression like

$$y = 3 + \log x^3 z,$$

to give an explicit expression for x in terms of the other variables.

The three main areas of mathematics we need are probability, linear algebra, and calculus. Most of the results you should know are summarised on the following cribsheet:
`http://homepages.inf.ed.ac.uk/imurray2/pub/cribsheet.pdf`

The rest of the document below gives some additional details and reading.

Probability

As stated in the cribsheet, MacKay's free textbook provides a terse introduction to probability, as does Murphy section 2.2, or Barber Chapter 1. Alternatively, Sharon Goldwater has a longer, more tutorial introduction:
`http://homepages.inf.ed.ac.uk/sgwater/math_tutorials.html`

You *must* know the sum and product rules of probability: their equations, what they mean, and how to apply them for discrete and real-valued variables.

Expectations, or averages of random quantities, are also important. I have provided detailed notes on those with the MLPR course notes.

Linear Algebra

An undergraduate linear algebra course will usually discuss abstract linear spaces and operators. This course largely focusses on concrete operations on matrices and vectors expressed as arrays of numbers, as we can explicitly compute in (for example) Matlab or Octave.

You need to be able to do basic algebraic manipulation of matrices and vectors, and know how matrix multiplication works. You should also have a geometric understanding of these operations, which can be relevant to understanding their application to machine learning. If you're unsure, please work through David Barber's tutorial:

`http://www.inf.ed.ac.uk/teaching/courses/mlpr/notes/mlpr-supplementary-maths.pdf`

A shortened version of this tutorial also appears as an appendix of his textbook.

You will *not* need to be able to numerically compute matrix inverses, determinants, or eigenvalues of matrices by hand for this course. You can safely skip those exercises!

There are many possible introductions to linear algebra. Another terse one is Chapter 2 of Goodfellow et al.'s Deep Learning textbook. A nice series of videos is 3blue1brown's Essence of Linear Algebra.

Differentiation

You should know how to differentiate algebraic expressions. Computer algebra systems can do this stuff for us, and I'll talk about automatic numerical differentiation later in the course.

However, in simple cases I still regularly differentiate expressions with pen and paper while doing research.

The cribsheet summarizes the basic results I expect you to know. If the rules don't make sense, you will need to consult an undergraduate level or advanced high-school level maths textbook, or a tutorial series such as those from Khan Academy.

Some students may not have seen or remember *partial derivatives*. For example:

$$\frac{\partial xy^2}{\partial x} = y^2, \quad \frac{\partial xy^2}{\partial y} = 2xy.$$

The curly ∂ simply means that you treat all other variables as constants when you are doing the differentiation.

Partial derivatives can be combined to create total derivatives. For example, imagine moving around the circumference of a circle by changing an angle θ. Your (x, y) position is given by $x = \cos\theta$ and $y = \sin\theta$. To compute the change in a function $f(\mathbf{x})$ due to an infinitesimal change $\mathrm{d}\theta$ in the angle, you can use the chain rule of differentiation:

$$\mathrm{d}f \;=\; \frac{\partial f}{\partial x}\,\mathrm{d}x + \frac{\partial f}{\partial y}\,\mathrm{d}y \qquad \text{or} \qquad \frac{\mathrm{d}f}{\mathrm{d}\theta} \;=\; \frac{\partial f}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}\theta} + \frac{\partial f}{\partial y}\frac{\mathrm{d}y}{\mathrm{d}\theta}.$$

In this case you could also substitute expressions for x and y, to find $f(\theta)$ and differentiate with respect to θ. You could try both methods to differentiate $f(x, y) = xy^2$ with respect to θ. You should get the same answer! The chain rule approach is needed in many machine learning settings.

You should be comfortable enough with both vectors and derivatives that you wouldn't find it intimidating to work with a vector containing derivatives. For example, if we want to find partial derivatives of a function $f(\mathbf{x})$ with respect to each element of a vector $\mathbf{x} = [x_1\ x_2]^\top$, then the vector $\nabla_{\mathbf{x}} f$ is defined as:

$$\nabla_{\mathbf{x}} f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix}.$$

We will do mathematics containing such expressions. For example, given the chain rule of differentiation above, you should be happy that

$$\frac{\mathrm{d}f}{\mathrm{d}\theta} = (\nabla_{\mathbf{x}} f)^\top \begin{bmatrix} \frac{\mathrm{d}x_1}{\mathrm{d}\theta} \\ \frac{\mathrm{d}x_2}{\mathrm{d}\theta} \end{bmatrix}.$$

The mathematics textbook I used and liked as an undergraduate covers this material: "Mathematical Methods for Physics and Engineering", Riley, Hobson, Bence. Although there are many other possible textbooks and online tutorials.

3blue1brown also have an Essence of Calculus playlist.

Integration

You should also know enough about integration to understand the sum rule and expectations for real-valued variables.

The two most common situations in machine learning are: 1) an integral is impossible, there is no closed form solution; or 2) an integral is easy, there is a trick to write down the answer.

For example, the Gaussian distribution (discussed in much more detail later in the notes) with mean μ and variance σ^2 has probability density function:

$$\mathcal{N}(x;\, \mu, \sigma^2) \;=\; \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2\sigma^2}(x-\mu)^2}.$$

We may have to compute various integrals involving this function. For example:

$$I = \int_{-\infty}^{\infty} (x + x^2)\, \mathcal{N}(x;\, \mu, \sigma^2)\, \mathrm{d}x.$$

We can express this integral in terms of expectations for which we already know the answers:

$$I \;=\; \mathbb{E}[x] + \mathbb{E}[x^2] \;=\; \mu \;+\; (\mathrm{var}[x] + \mathbb{E}[x]^2) \;=\; \mu + \sigma^2 + \mu^2.$$

Summary: there's no need to revise everything about integration covered in a calculus course. I learned about many tricks for solving integrals that I never use, such as clever trigonometric substitutions and contour integration. However, you do need to be comfortable enough with what integration is, and with probability theory, so that you can follow and produce mathematical arguments like those above.