# Machine Learning and Pattern Recognition, Tutorial Sheet Number 1 Answers

School of Informatics, University of Edinburgh, Instructor: Chris Williams

## The students have been given answers to these and there should be no need to discuss these in class unless time permits and there is demand

1. Probability. Lois knows that on average radio station RANDOM-FM plays 1 out of 4 of her requests. If she makes 3 requests, what is the probability that at least one request is played?

   **Probability**.

   Random variable $X$ denotes the "number of requests played". Then

   $$P(X \geq 1) = P(X = 1) + P(X = 2) + P(X = 3) = 1 - P(X = 0)$$
   $$P(X = 0) = (0.75)^3$$
   $$P(X \geq 1) = 0.5781$$

2. Let $A$ and $\boldsymbol{v}$ be defined as

   $$A = \begin{pmatrix} 1 & 1 & 3 \\ 1 & 1 & -3 \\ 3 & -3 & -3 \end{pmatrix} \qquad \boldsymbol{v} = \begin{pmatrix} 1 \\ -1 \\ -2 \end{pmatrix}$$

   Calculate $A\boldsymbol{v}$. Is $\boldsymbol{v}$ an eigenvector of $A$, and if so what is the corresponding eigenvalue?

   **Matrix operations, eigenvalues and eigenvectors**.

   If $A\boldsymbol{v} = \lambda\boldsymbol{v}$, then $\boldsymbol{v}$ is an eigenvector of $A$ and $\lambda$ is the associated *eigenvalue*. Solution $\boldsymbol{v}$ is an *eigenvector* of $A$ with $\lambda = -6$. Write out $A\boldsymbol{v}$. Show it is a multiple of $\boldsymbol{v}$.

3. A random vector $\boldsymbol{x}$ has zero mean a diagonal covariance

   $$E(\boldsymbol{x}\boldsymbol{x}^T) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

   where $E$ stands for expectation (or mean average) of a random variable. If $\boldsymbol{y} = A^T\boldsymbol{x}$ (using $A$ from Q1) what is the covariance of the resulting random vector $\boldsymbol{y}$: $E(\boldsymbol{y}\boldsymbol{y}^T)$? You may use the fact that expectation is linear: $E(R\boldsymbol{x}\boldsymbol{x}^T S) = RE(\boldsymbol{x}\boldsymbol{x}^T)S$. This shows how covariances change under linear transformations.

   Do this with vectors first to get used to handling vectors and matrices: $\boldsymbol{y} = A^T\boldsymbol{x}$ (explain the transpose), so $\boldsymbol{y}^T = \boldsymbol{x}^T A$ (explain the rules of transposition). So $\boldsymbol{y}\boldsymbol{y}^T = A^T\boldsymbol{x}\boldsymbol{x}^T A$. So $E(\boldsymbol{y}\boldsymbol{y}^T) =$

$E(A^T \boldsymbol{x}\boldsymbol{x}^T A) = A^T E(\boldsymbol{x}\boldsymbol{x}^T)A = A^T A$ (explain the identity matrix). Then compute $A^T A$ if you want to: `[11 -7 -9;-7 11 9; -9 9 27]`. Spend a little time demonstrating linear transforms (e.g. use a rectangle, change size, elongate in one direction, shear rectangle), and understand that different covariances are all just produced by taking identity matrix and using linear transformation. Hence different Gaussians are produced in the same way: all Gaussians are equivalent under linear transformation and displacement.

4. Find the partial derivatives of the function $f(x, y, z) = (x + 2y)^2 \sin(xy)$.

**Partial Derivatives**

$$\begin{aligned}
\frac{\partial f}{\partial x} &= y(x+2y)^2 \cos(xy) + 2(x+2y)\sin(xy) \\
\frac{\partial f}{\partial y} &= x(x+2y)^2 \cos(xy) + 4(x+2y)\sin(xy) \\
\frac{\partial f}{\partial z} &= 0
\end{aligned}$$

5. Let $x$ be a Gaussian random variable with mean $\mu$ and variance $\sigma^2$. What is the expected value of $2x^2$? Show what form the distribution of $2(\mathsf{x}-\mu)^2$ takes. Hint: the distribution of $x^2$ for a standard normal (N(0,1)) is chi-squared distributed.

Expectation:

$$I = \int_{-\infty}^{\infty} dx\ 2x^2 \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right).$$

Writing $x^2 = (x-\mu)^2 + 2x\mu - \mu^2$ we have

$$\begin{aligned}
I = 2\int_{-\infty}^{\infty} dx\ (x-\mu)^2 \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) \\
+ 4\mu \int_{-\infty}^{\infty} dx\ \frac{x}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) - 2\mu^2 \quad (1)
\end{aligned}$$

The first is the 2 times variance ($2\sigma^2$). The second is $4\mu\mu$ and the third is $-2\mu^2$. Putting all this together we get $2\sigma^2 + 2\mu^2$. Can work through this further (e.g. derive mean and variance) if desired.

Use the change of variables for a probability density. Remember that a density must integrate to one and so need to use the Jacobian (or derivative in 1D). Explain the Jacobian.

$$y = 2(x-\mu)^2 \text{ so } |dy/dx| = 4(x-\mu) = 4\sqrt{y/2} \text{ so } |dx/dy| = 1/\sqrt{8y}$$

But note that y(x) is double valued. Hence the density at $y$ is affected by two points $x$. Hence

$$P(y) = 2P(x)|dx/dy| = \frac{1}{\sqrt{4y\pi\sigma^2}} \exp\left(-\frac{y}{4\sigma^2}\right)$$

Plot this in Matlab so you know what it looks like.

.