Maths Quiz for Unsupervised Learning

GI18, Term 1

The purpose of this quiz is to help you and us assess any areas of background knowledge you may need to work on. **You do not need to turn it in**, but if you are not sure of any of your answers please speak with one of the TAs or instructors. If you cannot solve the majority of the questions, this course is probably not appropriate for you.

1. Consider two discrete random variables X and Y. Show that

$$P(X = x, Y = y) \leq P(X = x)$$

2. Given a list of numbers $x_1, x_2, \ldots x_n$ as input, write down pseudocode that will output the largest number in the list (obviously, don't rely on a `max` function, as found in matlab!).

3. You observe n pairs of data points $\{(x_1, y_1), \ldots, (x_n, y_n)\}$ and you want to fit the parameter w of a linear regression model of the form:

$$y = wx$$

   by minimising $\sum_{i=1}^{n}(y_i - wx_i)^2$. Show how you would find w.

4. This is a Gaussian density with mean μ and variance σ^2:

$$p(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right].$$

   What is the point x^* with the highest probability density? Back up your claim by finding the stationary point(s) of the density function.

   Now let $y = (x - \mu)^2$. What is the density function of y? What value of y has highest density? [Hint: it isn't 0.]

5. Let $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$ be $n \times 1$ column vectors and A be an $n \times n$ matrix. Assume that you know that $\mathbf{x}^\mathsf{T} A \mathbf{y} = 1$, where the superscript $^\mathsf{T}$ denotes the transpose operator. For each of the following statements answer **true**, **false**, or **dimensionally inconsistent** (if the two sides of the equation have different dimensionalities).

   $\mathbf{x}^\mathsf{T} A \mathbf{y} = 1$ implies that:

   (a) $\mathbf{y}^\mathsf{T} A \mathbf{x} = 1$

   (b) $A \mathbf{x}^\mathsf{T} \mathbf{y} = 1$

   (c) $\mathbf{y}^\mathsf{T} A^{-1} \mathbf{x} = 1$

   (d) $\mathbf{y}^\mathsf{T} A^\mathsf{T} \mathbf{x} = 1$

   (e) $\mathbf{x}^\mathsf{T} A(\mathbf{y} - \mathbf{z}) = 1 - \mathbf{x}^\mathsf{T} A \mathbf{z}$

   (f) $\mathsf{Trace}(A \mathbf{y} \mathbf{x}^\mathsf{T}) = 1$

6. The weather in London has probability p of being rainy on any given day. Assume the weather is independent across days. Let $X_i = 1$ if it is rainy on day i, and $X_i = 0$ if it is not rainy. Using the notation $E(X_i)$ to mean the "expected value of X_i" and $\mathrm{Var}(X_i)$ to mean the "variance of X_i", compute the values (in terms of p) of the following quantities:

   (a) $E(X_i)$

   (b) $E(X_i^2)$

   (c) $\mathrm{Var}(X_i)$

   (d) $E(\sum_{i=1}^{n} X_i)$

7. What is $\displaystyle\int_{\mu}^{\infty} dx\, e^{-(x-\mu)^2/2\sigma^2}$?

8. Let $A(x)$ be a $(n \times n)$-matrix-valued function of a scalar x.

   (a) What is the dimensionality of $\dfrac{dA}{dx}$?

   (b) What is the dimensionality of $\dfrac{d|A(x)|}{dx}$? (The vertical bars represent the determinant operator).

   (c) What is $\dfrac{d|A(x)|}{dx}$?

9. Show that for small x

$$\log(1 + x) \approx x - \frac{x^2}{2}.$$

   [Hint: consider using a Taylor series expansion]