STA414 Assignment #1

Due: Never. This is not for credit

Question 1 (Variance and covariance, 6 marks)

Let A and B be two independent random variables.

(a) Show that if A and B are independent then their covariance is zero.

(b) Letting a be a scalar constant, show that:

$$\mathbb{E}(A + aB) = \mathbb{E}(A) + a\mathbb{E}(B)$$
$$\text{var}(A + aB) = \text{var}(A) + a^2\text{var}(B)$$

Question 2 (Densities, 5 marks)

Each of the following questions is worth one mark. yes, as long as integration over domain is 1

(a) Is it possible for a probability density function (pdf) to take a value greater than 1?

(b) Suppose X is a univariate normally distributed random variable with a mean of 0 and variance of $1/16$. What is the pdf of X?

(c) State the value of this pdf at 0.

(d) State the probability that $X = 0$. =0 for point estimate

(e) Describe why the answers to (c) and (d) are different.

pdf is derivative of cdf, integral of pdf over set A equal probability

Question 3 (Calculus, 4 marks)

Answer the following, using vector notation in your responses. You are given that $\mathbf{z}, \mathbf{y} \in \mathbb{R}^n$ and $\mathbf{A} \in \mathbb{R}^{n\times n}$.

(a) What is the gradient w.r.t. $\mathbf{z}$ of $\mathbf{z}^T\mathbf{y}$? y

(b) What is the gradient w.r.t. $\mathbf{z}$ of $\mathbf{z}^T\mathbf{z}$? 2z

(c) What is the gradient w.r.t. $\mathbf{z}$ of $\mathbf{z}^T\mathbf{A}\mathbf{z}$? (A+A^T)z

(d) What is the gradient w.r.t. $\mathbf{z}$ of $\mathbf{A}\mathbf{z}$? A

Question 4 (Regression, 7 marks)

Let $\mathbf{X} \in \mathbb{R}^{n\times m}$ be a matrix such that $n \geq m$. Let $\mathbf{y} \in \mathbb{R}^n$ be a vector such that $\mathbf{y} \sim \mathcal{N}(\mathbf{X}\boldsymbol{\beta}, \sigma^2 I)$. Recall from the lectures that the MLE of $\boldsymbol{\beta}$ is given by

$$\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$$

Note Y = XB + epsilon
epsilon ~ N(0, sigma^2 I)

(a) Explain why it is necessary for $n \geq m$.

(b) For a given value of $\boldsymbol{\beta}$, give the expectation and covariance matrix of $\hat{\boldsymbol{\beta}}$.

(c) Assume that we have observed $\mathbf{y}$ and $\mathbf{X}$. Find the gradient of the log likelihood w.r.t. $\boldsymbol{\beta}$.

Question 5 (Ridge Regression, 4 marks) where variance is sigma^2

If we express our prior knowledge about $\boldsymbol{\beta}$ using a normal distribution, we can assume that $\boldsymbol{\beta} \sim \mathcal{N}(0, \tau^2 \mathbf{I})$. The MAP estimate of $\boldsymbol{\beta}$ given $\mathbf{y}$ in this context is

$$\hat{\beta}_{\text{MAP}} = (\mathbf{X}^T\mathbf{X} + \lambda \mathbf{I})^{-1}\mathbf{X}^T\mathbf{Y}$$

where $\lambda = \sigma^2/\tau^2$. Recall from the lectures that this is called *ridge regression*.

No, invertible regardless of rank

(a) In order to do ridge regression, is it necessary that $n \geq m$? Explain why or why not.

(b) There is an equivalent way to expressing the above equation for $\hat{\beta}_{\text{MAP}}$. Starting with the equation for $\hat{\beta}$ in Question 4, augment $\mathbf{X}$ by adding m rows to it in which most entries are zero. The nonzero entries are such that, for the ith additional row, the entry in the ith column is equal to $\sqrt{\lambda}$. Then add m corresponding entries to $\mathbf{y}$ that are all 0. Show that this is equivalent to calculating $\hat{\beta}_{\text{MAP}}$ above.

Question 6 (High dimensions, 4 marks)

A hypersphere is the generalization of the concept of a sphere, to arbitrary dimension (not just $d = 3$). Consider a d-dimensional hypersphere of radius r. The fraction of its hypervolume lying between values $r - c$ and r, where $0 < c < r$, is given by

$$f = 1 - \left(1 - \frac{c}{r}\right)^d.$$ fraction of volumne near surface

(a) For any fixed c value, f tends to 1 as $d \to \infty$. Show this numerically, with $c/r = 0.01$, for $d = 2$, 10, and 1000.

(0.0199, 0.095, 0.9999)

(b) Evaluate the fraction of the hypervolume which lies inside the radius $r/2$ for $d = 2$, 10, and 1000.

(0.25, 0.000097, 0.000000)

(c) The figure below shows points distributed according to the uniform distribution inside a circle. For uniformly distributed points inside a very high-dimensional hypersphere centred at the origin, select one of the following as correct (no explanation needed):

(a) Most points are found near the middle of the hypersphere (the origin),

(b) Most points are found along the axes,

(c) Most points are found close to the hypersphere's surface, or

(d) None of the above.