DatA414: Tutorial 1

Question 1

For the simple linear regression model

$$f(x; w_0, w_1) = w_0 + w_1 x$$

proove that $\frac{\partial^2 J}{\partial w_0^2} > 0$ for the squared loss J. Why is it important/useful to show this?

Question 2

Proove that:

$$\frac{\partial \mathbf{x}^\top \mathbf{A} \mathbf{x}}{\partial \mathbf{x}} = 2\mathbf{A}\mathbf{x}$$

if $\mathbf{A}$ is symmetrical.

Question 3

For L_2-regularised linear regression we have the following loss:

$$J_\lambda(\mathbf{w}) = \sum_{n=1}^{N} \left(y^{(n)} - f(\mathbf{x}^{(n)}; \mathbf{w}) \right)^2 + \lambda \sum_{k=1}^{K} w_k^2$$

Prove that the closed-form solution for the weights that minimise this loss is given by:

$$\hat{\mathbf{w}}_\lambda = (\mathbf{X}^\top \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^\top \mathbf{y}$$

In the exam, I could ask you to write the Python code to calculate these weights (I will be slightly lenient if you forget the exact function names, as long as I can see that you actually did and understood the practical).

Question 4

What is the difference between L_1 and L_2 regularisation?

Question 5

Suppose we have the following dataset:

n	$x^{(n)}$	$y^{(n)}$
1	1896	12.00
2	1900	11.00
3	1904	11.00
4	1906	11.20

We perform regression on this dataset using basis functions, with $\boldsymbol{\phi}(x) = \begin{bmatrix}1 & \phi_1(x) & \phi_2(x)\end{bmatrix}^\top$. We use radial basis functions (RBFs) for $\phi_1(x)$ and $\phi_2(x)$ centered at $c = 1896$ and $c = 1906$, respectively, both with a width parameter of $h = 10$. The multidimensional RBF is defined as:

$$\exp\left\{\frac{-(\mathbf{x}-\mathbf{c})^\top(\mathbf{x}-\mathbf{c})}{h^2}\right\}$$

Write out the basis function design matrix $\boldsymbol{\Phi}$ for the dataset above.

Question 6

Prove that:

- $\text{var}[x + y] = \text{var}[x] + \text{var}[y]$ if x and y are statistically independent
- $\text{var}[cx] = c^2\text{var}[x]$

You should be able to do this without writing out an integral.

Question 8

A discrete random variable has a Bernoulli distribution if it takes the value 1 with probability p and the value 0 with probability $1 - p$. I.e., it has a probability mass function:

$$P(x) = \begin{cases} p & x = 1, \\ 1-p & x = 0, \\ 0 & \text{otherwise,} \end{cases}$$

If we have N IID samples from a Bernoulli random variable $\{x^{(n)}\}_{n=1}^N$, what is the maximum likelihood estimate of the parameter p?

Question 9

What is the value of the integral

$$Z = \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}(u-\mu)^2}\ \mathrm{d}u$$

By considering $\int_{-\infty}^{\infty} \mathcal{N}(u; \mu, \sigma^2)\ \mathrm{d}u$, you should be able to write down the answer without doing any detailed calculations.