Machine Learning/Deep Learning Pre-quiz
AMSI Summer School 2023

Susan Wei, University of Melbourne

August 25, 2023

Problem 1 *Suppose we have a column vector x and a square matrix A. Let $f(x) = x^T Ax$. Show that*

$$\nabla f(x) = (A + A^T)x.$$

Problem 2 *Suppose that X is an $n \times m$ matrix, U is an $n \times r$ matrix, and V is an $r \times m$ matrix. Let $f(X, U, V) = ||X - UV||_2^2$. Show that*

$$\nabla_X f(X, U, V) = -2U^T (X - UV).$$

Problem 3 *Take the function $f(x, y) = \log(e^x + e^y)$. Show that the gradient of f at $(0, \log(2))$ is given by $[1/3, 2/3]$. Approximate f at $(\epsilon_1, \log(2) + \epsilon_2)$ using the gradient.*

Problem 4 *Suppose you have two random variables X, Y with joint density $p_{X,Y}(x, y) = 4xy$ for $x, y \in [0, 1]$ and zero otherwise. What is the covariance of X and Y?*

Problem 5 *You obtain a single observation which is the number 3 from the random variable which has density $\alpha \exp(-\alpha x)$ for some value $\alpha > 0$. What is the maximum likelihood estimate for α?*

Problem 6 *Show that the KL divergence $D(q||p) := E_q[\log q - \log p]$ is nonnegative for all distribution p and q.*

Problem 7 *For any random variable X that follows a probability distribution P with pdf p, the Shannon entropy is defined as $H(X) = -E_{x\sim P}[\log p(x)]$. We can similarly define the joint entropy $H(X, Y)$ of a pair of random variables (X, Y) as $H(X, Y) = -E_{(x,y)\sim P}[\log p_{X,Y}(x, y)]$. The conditional entropy $H(Y|X)$ is defined as $H(Y|X) = -E_{(x,y)\sim P}[\log p(y|x)]$.*

The mutual information of (X, Y), written $I(X, Y)$, measure how much information is shared between X and Y, loosely speaking. It is defined as

$$I(X, Y) = H(X, Y) - H(Y|X) - H(X|Y).$$

Test your intuition by confirming the following quantities are all equivalent to $I(X, Y)$:

- $H(X) - H(X|Y)$
- $H(Y) - H(Y|X)$
- $H(X) + H(Y) - H(X, Y)$