# Homework 1 - Math Review

Machine Learning - CSC 84020 - Prof. Rosenberg

Due February 18 at 2:00pm

Problem 1) (5 pts.) Prove that if two variables $x$ and $y$ are independent, their covariance is zero.

Problem 2) (10 pts.) Consider two variables $x$ and $y$ with joint distribution $p(x, y)$. Prove the following two results.

$$\begin{aligned} \mathbb{E}[x] &= \mathbb{E}_y[\mathbb{E}_x[x|y]] \\ \mathrm{var}[x] &= \mathbb{E}_y[\mathrm{var}_x[x|y]] + \mathrm{var}_y[\mathbb{E}_x[x|y]] \end{aligned}$$

Here $\mathbb{E}_x[x|y]$ denotes the expectation of $x$ under the conditional distribution $p(x|y)$, with a similar notation for the conditional variance.

Problem 3) Bayes Rule.

You are given three boxes, one Red, one Blue and one Green.

- The Red box contains 1 Orange, 4 Apples and 5 Cherries.
- The Blue box contains 7 Orange, 2 Apples and 1 Cherries.
- The Green box contains 3 Orange, 4 Apples and 3 Cherries.

3a. (5 pts.) Assuming an even prior distribution, what are the posterior probability that an Orange was drawn from a Red, Green, or Blue box?

3b. (5 pts.) Assume a prior distribution $\boldsymbol{\mu} = \begin{bmatrix} \mu_{Red} & \mu_{Blue} & \mu_{Green} \end{bmatrix}^T = \begin{bmatrix} 0.2 & 0.5 & 0.3 \end{bmatrix}^T$. what are the likelihoods of a Red, Green, and Boxes given the selection of a Cherry?

Problem 4) (10 pts.) Maximum likelihood estimates of Gaussian parameters.

Given set of data $\mathbf{x}$, identify the most likely values of the Gaussian parameters $\mu$ and $\sigma^2$.

The likelihood of $\mathbf{x}$ given Gaussian parameters, $\mu$ and $\sigma^2$, is calculated as follows.

$$p(\mathbf{x}|\mu, \sigma^2) = \prod_{i=0}^{N-1} \frac{1}{\sqrt{2\pi\sigma^2}} exp\left\{ -\frac{1}{2\sigma^2}(x_i - \mu)^2 \right\}$$

Since $\ln(x)$ is a monotonic function, identifying the parameters that maximize the likelihood is equivalent to identifying the parameters that maximize the log likelihood. Maximizing the log likelihood function will by much easier.

Calculate the maximum likelihood parameters by taking the partial derivative of the log likelihood function with respect to $\mu$ and $\sigma^2$.

Problem 5) Vector Calculus.

5a. (15 pts.)
Let,

$$\mathbf{x} = \begin{bmatrix} x_0 \\ x_1 \\ x_2 \end{bmatrix}$$

Evaluate the following

$$\nabla_{\mathbf{x}} f(\mathbf{x})$$

where

$$f(\mathbf{x}) = 2x_0^3 + \sin(x_0) + x_0 x_1^2 + \frac{\exp\{(x_1 - x_2)^2\}}{x_2}$$

5b. (15 pts.) Evaluate the following.

$$\frac{\partial}{\partial \mathbf{x}} \left( \mathbf{x}^T \mathbf{A} + (\mathbf{A}^T \mathbf{x})^T + (\mathbf{x}^T \mathbf{A}^{-1} \mathbf{x})(\mathbf{x}^T \mathbf{y}) \right)$$

Problem 6) Conjugate Priors (25pts.)

Consider a single Gaussian random variable $x$. Suppose that $\sigma^2$ is known. Estimate the mean $\mu$ given the a set of $N$ observations $\mathbf{x} = \{x_0, x_1, \ldots, x_{n-1}\}$.

Assume the following function to define the likelihood of $\mathbf{x}$ given $\mu$. Note that this is not normalized and is thus **not** a probability distribution function.

$$p(\mathbf{x}|\mu) \quad = \quad \prod_{i=1}^{N-1} p(x_n|\mu) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{n=1}^{N} (x_n - \mu)^2 \right\}$$

Also assume the following *prior* distribution for $\mu$.

$$p(\mu) = N(\mu|\mu_0, \sigma_0^2)$$

The posterior is given by the following

$$p(\mu|\mathbf{x}) \propto p(\mathbf{x}|\mu)p(\mu)$$

Show that

$$p(\mu|\mathbf{x}) = N(\mu|\mu_N, \sigma_N^2)$$

where

$$\begin{aligned}
\mu_N &= \frac{\sigma^2}{N\sigma_0^2 + \sigma^2}\mu_0 + \frac{N\sigma_0^2}{N\sigma_0^2 + \sigma^2}\mu_{ML} \\
\mu_{ML} &= \frac{1}{N}\sum_{n=0}^{N-1} x_n \\
\frac{1}{\sigma_N^2} &= \frac{1}{\sigma_0^2} + \frac{N}{\sigma^2}
\end{aligned}$$

Hint: you will need to "complete the square" as part of your algebraic manipulation to the gaussian form.