CS 182: Introduction to Machine Learning, Fall 2021
Homework 1

(Due on Monday, Oct. 11 at 11:59pm (CST))

Notice:

- Please submit your assignments via Gradescope. The entry code is KYJ626.
- Please make sure you select your answer to the corresponding question when submitting your assignments.
- Each person has a total of five days to be late without penalty for all the homeworks. Each late delivery less than one day will be counted as one day.

1. [20 points]

   (a) Given a set of observation pairs $\{(x_i, y_i)\}_{i=1}^N$, where $x_i, y_i \in \mathbb{R}$, $i = 1, 2, ..., N$. By assuming the linear model is a reasonable approximation, we consider to fit the model via the least squares method. Thus, our goal is to estimate the coefficients $\hat{\omega}_0$ and $\hat{\omega}_1$ to minimize the residual sum of squares (RSS),

   $$[\hat{\omega}_0,\ \hat{\omega}_1] = \underset{\omega_0,\ \omega_1}{\text{argmin}} \sum_{i=1}^{N} [y_i - (\omega_1 x_i + \omega_0)]^2. \tag{1}$$

   Please show that

   $$\begin{cases} \hat{\omega}_1 = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^N (x_i - \bar{x})^2}, \\ \hat{\omega}_0 = \bar{y} - \hat{\omega}_1 \bar{x}, \end{cases} \tag{2}$$

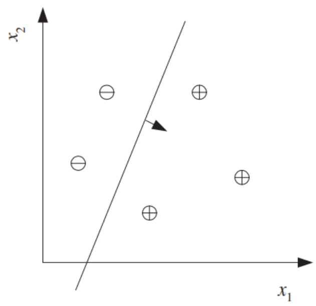
   where $\bar{x} = \frac{1}{N}\sum_{i=1}^N x_i$ and $\bar{y} = \frac{1}{N}\sum_{i=1}^N y_i$ denote the sample means. [10 points]

   (b) Assume now we want to classify the examples $\{x_i\}_{i=1}^N$, $x_i \in \mathbb{R}$, $i = 1, \cdots, N$ and the hypothesis class is

   $$\mathcal{H}(x) = \begin{cases} 1 & a \le x \le b, \quad a, b \in \mathbb{R}, a < b \\ 0 & otherwise. \end{cases} \tag{3}$$

   What is the VC dimension of $\mathcal{H}$ and why? (You need to show that if the VC dimension is k, k points can be shattered but $k+1$ points cannot. See P40 of Lecture 01.) [5 points]

   (c) Assume now the examples $\{\mathbf{x}_i\}_{i=1}^N$, $\mathbf{x}_i \in \mathbb{R}^2$, $i = 1, \cdots, N$ and the hypothesis class is the set of lines. What is the VC dimension of $\mathcal{H}$ and why? [5 points]



2. [20 points] Suppose we have a two-class recognition problem with ω_1 and ω_2. The $p(x|\omega_i)$ follows normal distribution such that

$$p(x|\omega_i) \sim \mathcal{N}(\mu_i, \sigma^2) \tag{4}$$

and $p(\omega_i)$ is known. Suppose we have $\mu_2 > \mu_1$.

(a) Write the discriminant functions $g_i(x)$ and the classfication rule. [10 points]

(b) Derive the boundary of the decision regions. [10 points]

3. [20 points] Given a set of observations $\{x_i\}_{i=1}^N$, where $x_i \in \mathbb{R}$, $i = 1, 2, ..., N$. Assume $\{x_i\}_{i=1}^N \sim \mathcal{N}(\theta, \sigma^2)$ and $\theta \sim \mathcal{N}(\theta_0, \sigma_0^2)$, where σ, θ_0 and σ_0 are known constants.

   (a) Derive the MLE of θ. [6 points]

   (b) Derive the MAP of θ. [7 points]

   (c) Derive the Bayes' estimator of θ. [7 points]

4. [20 points] Given a set of observation pairs $\{(x_i, y_i)\}_{i=1}^N$, where $x_i, y_i \in \mathbb{R}$, $i = 1, 2, ..., N$. By assuming the polynomial model is a reasonable approximation, we consider to fit the model via the least squares estimate. Consider the polynomial regression function of order k:

$$g(x_i|\omega_0, \cdots, \omega_k) = \sum_{j=0}^{k} \omega_j x_i^j. \tag{5}$$

Define $\boldsymbol{\omega} = [\omega_0, \cdots, \omega_k]^T$. Show that the least squares estimate of $\boldsymbol{\omega}$ (assuming that $\mathbf{A}^T\mathbf{A}$ is invertible) is

$$\hat{\boldsymbol{\omega}} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{y}, \tag{6}$$

where

$$\mathbf{A} = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^k \\ 1 & x_2 & x_2^2 & \cdots & x_2^k \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_N & x_N^2 & \cdots & x_N^k \end{bmatrix}, \tag{7}$$

$$\hat{\boldsymbol{\omega}} = [\hat{\omega}_0, \cdots, \hat{\omega}_k]^T \tag{8}$$

and

$$\mathbf{y} = [y_1, \cdots, y_N]^T. \tag{9}$$

5. [20 points] Given a set of observations $\{x_i\}_{i=1}^{N}$ that are drawn i.i.d. from a Poisson distribution

$$P(x \mid \lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$$

with parameter $\lambda > 0$.

(a) Derive the MLE of λ and determine whether it is unbiased or not. [10 points]

(b) Derive the MLE of $\eta = e^{-2\lambda}$ and determine whether it is unbiased or not. [10 points]

(You need to give the bias if the estimator is biased.)