ML HW-1

Thursday, September 22, 2022 1:23 PM

HW-1

CS 559: Machine Learning

Homework Assignment 1

Due Date: Tuesday 6:30 PM, September 27, 2022 **Total: 100 points**

Problem 1. Probability (10 points)

Suppose that there are three sessions for machine learning course, including Session 1 (S1), Session 2 (S2), and Session 3 (S2), with students coming from different majors, including Computer Science (CS), Statistics (STAT), and Management (MG). The number of students in each session is shown in the following table.

Major	Session 1	Session 2	Session 3
CS	6	10	6
STAT	8	10	6
MG	6	0	8

If a session is chosen at random with probability $p(S1) = 0.2, p(S2) = 0.2, p(S3) = 0.6$, and a student in the selected session is selected to give a presentation about machine learning (with equal probability of selecting any of the students in the session), please find the following probabilities.

1) The probability of select a student majored in CS. $22/60 = 0.3666 = 36.6\%$
2) If we know that the selected student is from STAT, what is the probability that the student comes from Session 3?

$P(STAT) = 24/60$

$P(S3) = 0.6$

$P(STAT|S3) = P(S3|STAT) \cdot P(STAT)$

$= 0.6(0.4)$

$\boxed{= 0.24}$

Problem 2. Maximum Likelihood (ML) Estimation (15 points)

Suppose that we randomly select 10 students in our class and collect their weights (in pounds) as follows.

112, 120, 131, 126, 145, 158, 157, 136, 148, 176

Assume that their weights are **normally distributed with unknow mean μ and variance σ^2**. Based on the assumption, please answer the following questions.

1) Give the probability of the data set given the two parameters, i.e. the likelihood function.
2) Derive and calculate the solution for both μ and σ^2 using Maximum Likelihood Estimation. Please provide the detailed steps about getting the log-likelihood and calculating the derivative with respect to both parameters.

Poisson

$$f(x_i; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right)}$$

$$1.\ L(\mu, \sigma) = \sigma^{-n} (2\pi)^{-n/2} e^{\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right]}$$

$$2.\ \log L(\mu, \theta) = -n \log(\theta) - \frac{n}{2}\log(2\pi) - \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}$$

$$\frac{dl(\sigma)}{d(\sigma)} = -\frac{n}{\sigma} - 0 - \frac{2\sigma^2 0 - \sum_{i=1}^{n}(x_i-\mu)^2 4\sigma}{2\sigma^{2^2}}$$

$$= -\frac{n}{\sigma} + \frac{\sum_{i=1}^{n}(x_i-\mu)^2 4\sigma}{2\sigma^4} = 0$$

$$\frac{dl(\mu)}{d(\mu)} L(\mu, \theta) = 0 - 0 - \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$= \frac{1}{10}\sum_{i=1}^{10}(112+120+131+126+145+158+157+136+148+176)$$

$$mle = \mu = 140.9$$

Problem 3. Maximum Likelihood (ML) Estimation (15 points)

In problem 2, we obtained the ML estimate for both parameters in normal distribution. In this problem, please follow the similar procedure to obtain the ML estimate for the parameters in the probability mass function of the discrete random variables.

Suppose that X is a discrete random variable that follows the probability mass function provided below. Where the parameter q satisfies $0 \le q \le 1$.

X	1	2	3	4
P(X)	$\frac{2q}{3}$	$\frac{q}{3}$	$\frac{2(1-q)}{3}$	$\frac{1-q}{3}$

There are 10 independent observations obtained from this distribution: (4, 1, 3, 2, 4, 3, 2, 1, 3, 2). Based on the assumption, please answer the following questions.

1) Give the probability of the data set given the parameter q, i.e. the likelihood function.
2) Derive and calculate the Maximum Likelihood estimate of q. Please provide the detailed steps about getting the log-likelihood and calculating the derivative with respect to q.

$$L(q) = P(x=4)P(x=1)P(x=3)P(x=2) P(x=4)P(x=3)P(x=2)P(x=1) P(x=3)P(x=2)$$

$$L(q) = \left(\frac{1-q}{3}\right)^2 \left(\frac{2(1-q)}{3}\right)^3 \left(\frac{q}{3}\right)^3 \left(\frac{2q}{3}\right)^2$$

$$\log L(q) = 2\left(\log\frac{1}{3} + \log(1-q)\right) + 3\left(\log\frac{2}{3} + \log(1-q)\right) + 3\left(\log\left(\frac{1}{3}\right) + \log(q)\right) + 2\left(\log\left(\frac{2}{3}\right) + \log(q)\right)$$

$$= C + 5\log(1-q) + 5\log(q)$$

$$\frac{dl(q)}{d(q)} = -\frac{5}{1-q} + \frac{5}{q} = 0 \text{ so MLE } q = 0.5$$

Problem 4. Maximum Posterior (MAP) Estimation (20 points)

Given N input values $\mathrm{x} = (x_1, \dots, x_N)^T$ and their corresponding target values $\mathrm{y} = (y_1, \dots, y_N)^T$, we estimate the target using function $f(x, w)$ which is a polynomial curve. Assuming the target variables are drawn from Gaussian distribution:

$$p(y|x, w, \beta) = \mathcal{N}(y|f(x, w), \beta^{-1})$$

and a prior Gaussian distribution for w:

$$p(w|\alpha) = \left(\frac{\alpha}{2\pi}\right)^{(M+1)/2} \exp\left(-\frac{\alpha}{2} w^T w\right)$$

Here, M+1 is the total number of parameters in w. Please **use Bayes' theorem** to prove that maximum posterior (MAP) is equivalent to minimizing the regularized sum-of-squares error function. Note that the posterior distribution of w is $p(w|x, y, \alpha, \beta)$. You can refer to the slides about MAP in Lecture 1 and page 30 in the book: "Bishop, Pattern Recognition and Machine Learning, Springer, 2006".

$$P(Y|X, w, \beta) = N(Y|f(X, w), \beta^{-1})$$

$$P(w|X, Y, a, \beta) = P(Y|X, w, \beta) P(w|a)$$

Take the negative logarithm

$$\log P(w|X, Y, a, \beta) = -\frac{\beta}{2}\sum_{n=1}^{n} f(x_n, w) - y_n)^2 + \frac{N}{2}\log\beta - \frac{N}{2}\log(2\pi) - \frac{N}{2}\log\beta + \frac{N}{2}\log(2\pi) + \frac{a}{2} w^T w$$

$$= \frac{\beta}{2}\sum_{n=1}^{N} f(x_n, w) - y_n)^2 + \frac{a}{2} w^T w$$

$$= \frac{\beta}{2}\sum_{n=1}^{N} \{f(x_n, w) - y_n)^2 + \frac{a}{2} w^T w$$

SoS $E_p(w) = \frac{1}{2}\sum_{n=1}^{n} (y(x_n, w) - t_n)^2$

Problem 5. Linear Regression (40 points)

The Energy Efficiency Dataset (ENB2012_data.xlsx) in this problem is a public dataset from UCI (https://archive.ics.uci.edu/ml/datasets/Energy+efficiency). In this data, the energy analysis is performed using 12 different building shapes simulated in Ecotect. They are different in several aspects, such as including the glazing area and the orientation. In the dataset, there are 768 samples and 8 features. In this problem, the goal is to **predict the cooling load (y2 in the datasets)**. The dataset is provided on Canvas.

Suppose we consider the following three models to predict the cooling load on this dataset, including 1) Lasso regression, 2) Ridge regression, and 3) Elastic Net regression. Please use 5-fold cross validation to compare their performance based on mean squared error (MSE). To implement these models, you can use python and the sklearn packages. Please submit the code along with your solutions to get full points.

✓