Introduction to Machine Learning
Homework 1

March 8, 2019

Academic integrity

Our lesson cares much more on academic integrity. No matter who should do our utmost to handle the establishment of academic integrity standard including the host teacher and assistants of this lesson. We hope you will have the same faith with us.

(1) Discussion between students is allowing. The work named by yourself must be completed by your own hands. Any kind of Copying from existing documents will be seen as illegal.

(2) Any kind of Copying from other people's fruits of labour(Publication or Internet documents) will be accused of plagiarism. The score of plagiarists will be canceled. Please mark the authors if you cited any public documents of them;

(3) Highly resemble homework will be seen as Coping. No matter who you are, the one who copy or the one who is copied, both of your score will be canceled. Please protect your homework not to be copied by others actively.

Homework submission notes

(1) Please follow the submission methods on the website;

(2) If you are not follow the methods or your submission format are not correct. we will deduct some score of your homework;

(3) Unless some special cases(such as illness), the submission over deadline will not be accepted and your score will be set as zero.

1   [20pts] Basic Probability and Statistics

The probability distribution of random variable X follows:

$$f_X(x) = \begin{cases} \frac{1}{2} & 0 < x < 1; \\ \frac{1}{6} & 2 < x < 5; \\ 0 & \text{otherwise.} \end{cases} \tag{1.1}$$

(1) [5pts] Please give the cumulative distribution function $F_X(x)$ for X;

(2) [5pts] Define random variable Y as $Y = 1/(X^2)$, please give the probability density function $f_Y(y)$ for Y;

(3) [10pts] For some random non-negative random variable Z, please prove the following two formulations are equivalent:

$$\mathbb{E}[Z] = \int_{z=0}^{\infty} z f(z) \mathrm{d}z, \tag{1.2}$$

$$\mathbb{E}[Z] = \int_{z=0}^{\infty} \Pr[Z \geq z] \mathrm{d}z, \tag{1.3}$$

Meantime, please calculate the expectation of random variable X and Y by these two expectation formulations to verify your proof.

2   [20pts] Strong Convexity

Let $D \in \mathbb{R}^2$ be a finite set. Define a function $E : \mathbb{R}^3 \to \mathbb{R}$ by

$$E(a, b, c) = \sum_{x \in \mathcal{D}} (ax_1^2 + bx_1 + c - x_2)^2. \tag{2.1}$$

(1) [10pts] Show that E is convex.

(2) [10pts] Does there exist a set D such that E is strongly convex? Proof or a counterexample.

3   [20pts] Transition Probability Matrix

Suppose x_k is the fraction of NJU students who prefer course A at year k. The remaining fraction $y_k = 1 - x_k$ prefers course B.

At year $k + 1$, $\frac{1}{5}$ of those who prefer course A change their mind. Also at the same year, $\frac{1}{10}$ of those who prefer course B change their mind (possibly after taking the problem 3 last year).

Create the matrix P to give $[x_{k+1} \quad y_{k+1}]^\top = P[x_k \quad y_k]^\top$ and find the limit of $P^k[1 \quad 0]^\top$ as $k \to \infty$.

4 [20pts] Hypothesis Testing

Yesterday, a student was caught by the teacher when tossing a coin in class. The teacher is very nice and did not want to make things difficult. S(he) wished the student to determine *if the coin is biased for heads* with $\alpha = 0.05$.

Also, according to the student's desk mate, the coin was tossed for 50 times and it got 35 heads.

(1) [10pts] Show all calculate and rules (hint: using z-test).

(2) [10pts] Calculate the p-value and interpret it.

5 [20pts] Performance Measures

We have a set of samples that we wish to classify in one of two classes and a ground truth class of each sample (denoted as 0 and 1). For each example a classifier gives us a score (score closer to 0 means class 0, score closer to 1 means class 1). Below are the results of two classifiers (C_1 and C_2) for 8 samples, their ground truth values (y) and the score values for both classifiers (y_{C_1} and y_{C_2}).

y	1	0	1	1	1	0	0	0
y_{C_1}	0.5	0.3	0.6	0.22	0.4	0.51	0.2	0.33
y_{C_2}	0.04	0.1	0.68	0.22	0.4	0.11	0.8	0.53

(1) [8pts] For the example above calculate and draw the ROC curves for classifier C_1 and C_2. Also calculate the area under the curve (AUC) for both classifiers.

(2) [8pts] For the classifier C_1 select a decision threshold $th_1 = 0.33$ which means that C_1 classifies a sample as class 1, if its score $y_{C_1} > th_1$, otherwise it classifies it as class 0. Use it to calculate the confusion matrix and the F_1 score. Do the same thing for the classifier C_2 using a threshold value $th_2 = 0.1$.

(3) [4pts] Prove Eq.(2.22) in Page 35. (AUC $= 1 - \ell_{rank}$).

6 [Bonus 10pts]Expected Prediction Error

For least squares linear regression problem, we assume our linear model as:

$$y = x^T \beta + \epsilon, \tag{6.1}$$

where ϵ is noise and follows $\epsilon \sim N(0, \sigma^2)$. Note the instance feature of training data $\mathcal{D}$ as $\boldsymbol{X} \in \mathbb{R}^{p \times m}$ and note the label as $\boldsymbol{Y} \in \mathbb{R}^n$, where n is the number of instance and p is the feature dimension. So the estimation of model parameter is:

$$\hat{\beta} = (\boldsymbol{X}\boldsymbol{X}^T)^{-1}\boldsymbol{X}\boldsymbol{Y}. \tag{6.2}$$

For some given test instance x_0, please proof the expected prediction error $\mathbf{EPE}(x_0)$ follows:

$$\mathbf{EPE}(x_0) = \sigma^2 + \mathbb{E}_{\mathcal{D}}[x_0^T(\boldsymbol{X}\boldsymbol{X}^T)^{-1}x_0\sigma^2]. \tag{6.3}$$

Please give the steps and details of your proof.(Hint: $\mathbf{EPE}(x_0) = \mathbb{E}_{y_0|x_0}\mathbb{E}_{\mathcal{D}}[(y_0 - \hat{y}_0)^2]$, you can also refer to the proof progress of variance-bias decomposition on the page 45 of our reference book)