机器学习导论

姓名：殷天润　　**学号：**171240565

2019 年 3 月 17 日

请独立完成作业，不得抄袭。
若参考了其它资料，请给出引用。
鼓励讨论，但需独立书写解题过程。

第一部分　作业

☑ Problem 1 (ML problem 1)

1. Please give the cumulative distribution function F$_X$(x) for X;
2. Define random variable Y as Y = $1/(X^2)$, please give the proba bility density function f$_Y$ (y) for Y ;
3. A proof problemThe 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) [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:

$$\mathbb{E}[Z] = \int_{z=0}^{\infty} z f(z) \mathrm{d}z, \tag{2}$$

$$\mathbb{E}[Z] = \int_{z=0}^{\infty} \Pr[Z \geq z] \mathrm{d}z, \tag{3}$$

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

Solution

1.

$$F_X(x)=\begin{cases} 0 & -\infty < x \le 0 \\ \frac{1}{2}x & 0 < x \le 1 \\ \frac{1}{2} & 1 < x \le 2 \\ \frac{1}{6}+\frac{1}{6}x & 2 < x \le 5 \\ 1 & otherwise. \end{cases}$$

2. $F_Y(Y \le y) = P(\frac{1}{x^2} \le y) = P(-\frac{1}{\sqrt{y}} \le x \le \frac{1}{\sqrt{y}}) = 1 - F_X(\frac{1}{\sqrt{y}})$

   Then, $f_Y(y) = F\prime_Y(y) = \frac{1}{2}y^{-\frac{3}{2}}P_x(\sqrt{\frac{1}{y}})$.

   Therefore,

$$f_Y(y)=\begin{cases} \frac{1}{12}y^{-\frac{3}{2}} & \frac{1}{25} < y \le \frac{1}{4} \\ \frac{1}{4}y^{-\frac{3}{2}} & 1 \le y < \infty \\ 0 & otherwise. \end{cases}$$

3. $\mathbb{E}[X] = \int_{z=0}^{\infty} \Pr[Z \ge z]\mathrm{d}z = \int_{z=0}^{\infty}\int_{t=0}^{t=z} 1dtf(z)dz = \int_{z=0}^{\infty} zf(z)\mathrm{d}z$

   (a) i. $\mathbb{E}[X] = \frac{1}{2}*(1-0)*\frac{1}{2}+\frac{1}{2}*(25-4)*\frac{1}{6} = \frac{1}{4}+\frac{7}{4} = 2$

   ii. $\mathbb{E}[X] = \int_{x=0}^{\infty} \Pr[X \ge x] = \int_{x=0}^{\infty}(1-Pr[X \le x]) = \int_{x=0}^{\infty} 1 - \mathrm{F_X x}$

   Therefore, $\mathbb{E}[X] = \frac{3}{4}+\frac{1}{2}+\frac{3}{4} = 2$

   (b) i. $\mathbb{E}[Y] = (\sum k_i)\int_{z=0}^{\infty}\sqrt{y} = +\infty$($k_i$为某个正的常数,不需要关心)

   ii. 注意到$F_X(x) = \frac{1}{2}x \quad (x \in (0,1]))$,并且$F_Y = 1 - F_X(\frac{1}{\sqrt{y}})$,所以$F_Y$不收敛,$\mathbb{E}[Y] = +\infty$

Problem 2 (ML problem 2)

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

$$E(a,b,c) = \sum_{x \in \mathcal{D}}(ax_1^2 + bx_1 + c - x_2)^2. \tag{4}$$

(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.

Solution

1. 准备使用黑塞阵来证明;

$$H_x = \sum_{x_1 \in \mathcal{D}}\begin{bmatrix} 2*x_1^4 & 2*x_1^3 & 2*x_1^2 \\ 2*x_1^3 & 2*x_1^2 & 2*x_1 \\ 2*x_1^2 & 2*x_1 & 2 \end{bmatrix},$$

特征值:对于求和中的某一个x_1,

$$|\lambda E - H_x| = \begin{vmatrix} x_1^4-\lambda & x_1^3 & x_1^2 \\ x_1^3 & x_1^2-\lambda & x_1 \\ x_1^2 & x_1 & 1-\lambda \end{vmatrix}$$

特征方程为(省略计算步骤):$\lambda^2(x^4+x^2+1-\lambda)=0$,$\lambda_1=\lambda_2=0, \lambda_3=1+x^2+x^4>0$ 因此黑塞矩阵中作为求和必定大于等于0,黑塞矩阵半正定,E是convex的。

2. 由第一题的作答,黑塞矩阵必有两个为0的特征值,因此它不是正定但是是半正定的,因此不存在

Problem 3 (ML problem 3)

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 $k \to \infty$.

Solution

1. $x_{k+1} = -\frac{1}{5}x_k + \frac{1}{10}y_k$,$y_{k+1} = \frac{1}{5}x_k - \frac{1}{10}y_k$
   所以

$$P = \begin{bmatrix} -\frac{1}{5} & \frac{1}{10} \\ \frac{1}{5} & -\frac{1}{10} \end{bmatrix},$$

计算特征值并对角化:

$$\begin{bmatrix} 0 & 1 \\ 0 & -1 \end{bmatrix}^{-1} \begin{bmatrix} 0 & 0 \\ 0 & -\frac{3}{10} \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 0 & -1 \end{bmatrix}$$

因此

$$P^k[1 \quad 0]^T = \begin{bmatrix} 0 & 1 \\ 0 & -1 \end{bmatrix}^{-1} \begin{bmatrix} 0 & 0 \\ 0 & (-\frac{3}{10})^k \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

当k$\to \infty$的时候,上面的式子趋向于0

Problem 4 (ML problem 4)

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.

Solution

1. 假设:

$$H_0 : \mu_0 = 0.5$$

因为$\alpha = 0.05$,查表得到$\mu_{\frac{\alpha}{2}} = 1.96$;

数据量较小,我在excel里面手动用$\sigma = \sqrt{\frac{1}{n+1} * \sum(x_i - 0.5)^2}$的公式算到了$\sigma$约为0.505076

计算:

$$|\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}| = |\frac{0.7 - 0.5}{0.505076272/7.071067811865}| \approx 2.8 > 1.96$$

所以落在了拒绝域,不太可能发生。(这硬币应该不均匀)

2. p-value=P$\{Z \geq 2.80\}$$\approx 1 - 0.9974 = 0.0026$;

这意味着硬币均匀重量的概率大约是0.0026,概率非常小

---

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 $= 1 - \ell_{rank}$).

Solution

1. 我用python的sklearn库做了ROC曲线,纵轴是TPR,横轴是FPR,如下面的图

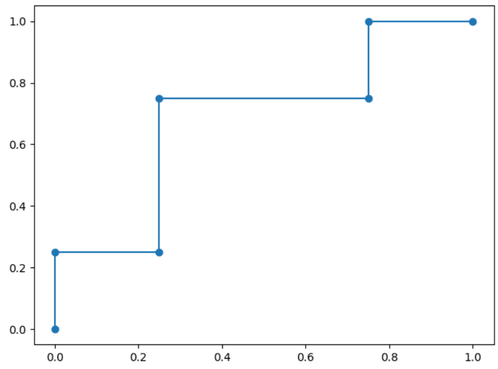

Figure 1: ROC figure of C1

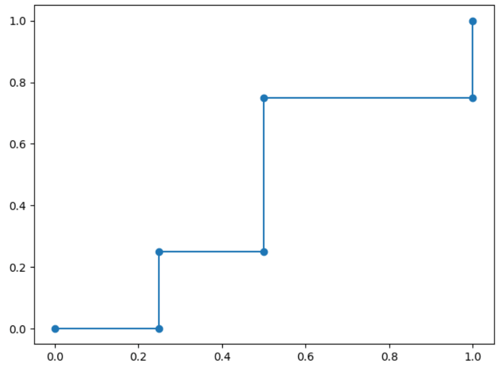

Figure 2: ROC figure of C2

其中,C1的AUC的值为0.6875,C2的AUC值为0.4375

2. C1的混淆矩阵

$$\begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}$$

$$P = 3/(3+1) = 0.75, R = 3/(3+3) = 0.5;$$

$$F1 = 2 * 0.75 * 0.5/(0.75+0.5) = 0.6;$$

C2的混淆矩阵

$$\begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix}$$

$$P = 2/(2+2) = 0.5, R = 2/(2+2) = 0.5;$$

$$F1 = 2 * 0.5 * 0.5/(0.5+0.5) = 0.5$$

Problem 6 (TC 29.4-2)

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

$$y = x^T\beta + \epsilon, \tag{5}$$

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:

$$\hat{\beta} = (\boldsymbol{X}\boldsymbol{X}^T)^{-1}\boldsymbol{X}\boldsymbol{Y}. \tag{6}$$

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

$$\mathbf{EPE}(x_0) = \sigma^2 + \mathbb{E}_{\mathcal{D}}[x_0^T(\boldsymbol{X}\boldsymbol{X}^T)^{-1}x_0\sigma^2]. \tag{7}$$

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)

Solution
