Machine Learning*
Homework LaTeX

* Teacher: ... TA: ...

1^{st} 张逸凯 171840708

Department of Computer Science and Technology

Nanjing University

zykhelloha@gmail.com

1. HOMEWORK I

A. [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) [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.

*Thank you teacher and TAs for your correction.

My solution:

(1):

$$F_X(x) = \begin{cases} 0 & x \leq 0; \\ \int_{-\infty}^{x} f_X(x) = \frac{x}{2} & 0 < x < 1; \\ \int_{-\infty}^{x} f_X(x) = \int_0^1 f_X(x) = \frac{1}{2} & 1 \leq x \leq 2 \\ \frac{1}{2} + \frac{x}{6} - \frac{1}{3} = \frac{x+1}{6} & 2 < x < 5; \\ 1 & x \geq 0 \end{cases}$$

(2):

Y 的可能取值范围: $(0, +\infty]$;

$y \leq 0$ 时,

$$F_Y(y) = P(Y \leq y) = 0;$$

$y > 0$ 时,

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

求导得:

$$p_Y(y) = F_Y'(y) = \begin{cases} \frac{1}{2y^{\frac{3}{2}}} \left(p_X(\frac{1}{\sqrt{y}}) - p_X(-\frac{1}{\sqrt{y}}) \right) & y > 0 \\ 0 & y \leq 0 \end{cases}$$

(3):

i. 由定义

$$\begin{aligned} &\mathbb{E}[Z] = \sum_{i=1}^{+\infty} z_i P(Z = z_i) \\ &= \lim_{\Delta z \to 0} \sum_{i=1}^{+\infty} z_i (F_Z(z_i + \Delta z) - F_Z(z_i)) \\ &= \lim_{\Delta z \to 0} \sum_{i=1}^{+\infty} z_i (\int_{z_i}^{z_i + \Delta z} f(z_i) \mathrm{d} z_i) \\ &= \lim_{\Delta z \to 0} \sum_{i=1}^{+\infty} \Delta z \, z_i f(z_i) \\ &= \int_{-\infty}^{\infty} z f(z) \mathrm{d} z \\ &= \int_{z=0}^{\infty} z f(z) \mathrm{d} z \end{aligned}$$

ii.

$$E(Z) = \int_0^{+\infty} (1-F(z))dz$$

$$0 \leq zF(-z) \leq \int_{-\infty}^{-z} |x| dF(x) \quad (\forall z>0)$$

$$\Rightarrow \lim_{z\to+\infty} zF(-z) \to 0 \Rightarrow \lim_{z\to-\infty} zF(z) = 0$$

同理 $0 \leq z[1-F(z)] \leq \int_0^{+\infty} x\, d(1-F(x))$

$$\Rightarrow \lim_{z\to+\infty} z[1-F(z)] = 0$$

$$\begin{aligned} E(Z) &= \int_0^{+\infty} z\, dF(z) \\ &= -\int_0^{+\infty} z\, d[1-F(z)] \\ &= -z(1-F(z))\Big|_0^{+\infty} + \int_0^{+\infty} (1-F(z))dz \\ &= \int_0^{+\infty} (1-F(z))dz \end{aligned}$$

得证.

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

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

(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.

A. [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 $k \to \infty$.

B. [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.

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

My solution:

(1):

注: 根据学习器预测结果对样例进行排序, 排在前面的是学习器认为"最可能"是正例的样本. 通过$x + \frac{1}{m^-}$等步骤更新点.

1	0	1	1	0	0	1	0	← 真实
0.6	0.51	0.5	0.4	0.33	0.3	0.22	0.2	← 预测

ROC

$(0,0)$

1 $(0, \frac{1}{4})$

0 $(\frac{1}{4}, \frac{1}{4})$

1 $(\frac{1}{4}, \frac{1}{2})$

1 $(\frac{1}{4}, \frac{3}{4})$

0 $(\frac{1}{2}, \frac{3}{4})$

0 $(\frac{3}{4}, \frac{3}{4})$

1 $(\frac{3}{4}, 1)$

0 $(1, 1)$

由AUC公式:

$$AUC = \frac{1}{2} \sum_{i=1}^{m-1} (x_{i+1} - x_i)(y_i + y_{i+1})$$

梯形, 上底加下底的和乘高除2.

(2):

(3):

$$l_{\text{rank}} = \frac{1}{m^+ m^-} \sum_{x^+ \in D^+} \sum_{x^- \in D^-} \left(\mathbb{I}\left(f\left(x^+\right) < f\left(x^-\right)\right) + \frac{1}{2} \mathbb{I}\left(f\left(x^+\right) = f\left(x^-\right)\right) \right)$$

注意到在画ROC curve的时候x轴$step_x = \frac{1}{m^-}$, y轴$step_y = \frac{1}{m^+}$.

所以xy平面可以被划分成面积是$step_x \times step_y$的小方块构成的, 由ROC curve绘制过程可以发现从一维的角度, 曲线上每一个平行于y轴的线段代表一个正例(即$x^+ \in D^+$), 同理平行于x轴的线段代表$x^- \in D^-$.

注意如果有多个真实分别是正例(s个)和假例(t个)的预测值一样, 这样降低分类阈值的时候会多个出现, 取下一个点就是:

$$\left(x+\frac{t}{m^{-}}, y+\frac{s}{m^{+}}\right)$$

所以其实ROC curve还会出现斜线, 不过问题不大, 因为我们可以发现这对书上AUC计算没有影响, 就变成计算梯形面积了.

从二维平面的角度$\sum_{x^{+}\in D^{+}}\sum_{x^{-}\in D^{-}}$ 代表对xy平面的$m^{+}\times m^{-}$进行遍历.

$f\left(x^{+}\right)<f\left(x^{-}\right)$ 代表通过遍历所有反样例来统计预测值大于x_i^{+}的预测值的反样例个数，也即该线段左边和下边的**平行于x轴线段**个数, 加上**倾斜线段**对应的反样例个数(即在x轴上投影有多少个$step_x$).

综上所述, 可以发现

$$l_{\text{rank}}+AUC=1$$

D. [Bonus 10pts]Expected Prediction Error

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 **EPE**(x_0) follows:

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

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