

机器学习导论_作业1

1 Basic concepts

1.1 Probability

解:

$$P(D|T) = \frac{P(D)P(T|D)}{P(T|D) \cdot P(D) + P(T|\neg D) \cdot P(\neg D)} = \frac{0.01 \times 0.98}{0.98 \times 0.01 + 0.10 \times (1 - 0.01)} = 0.09$$

故Bob有0.1088的概率确实患有该疾病。

1.2 Maximum likelihood estimation

解: 抛10次该硬币, 有8次正面朝上记为事件A, 那么有

$$P(A) = p^8(1-p)^2$$

记函数 $f(p) = p^8(1-p)^2, 0 < p < 1$

令 $f'(p) = 2p^7(1-p)(4-5p) = 0, 0 < p < 1$

得 $p = 0.8$

由于 $f'(p) > 0, (0 < p < 0.8)$ 且 $f'(p) < 0, (0.8 < p < 1)$, 那么在 $p = 0.8$ 时, $f(p)$ 取得最大值。

也即 $p = 0.8$ 时事件A发生的概率最大。故根据MLE, p 的估计值为0.8。

1.3 Performance measure

1) 根据 C_1, C_2 给出的预测结果分别对样本排序, 各样本划分为正例, 若当前为真正例, 则对应坐标为 $(x, y + 0.2)$, 若当前为假正例, 则对应坐标为 $(x + 1/3, y)$ 。

① C_1 :

y	yC_1	(x_i, y_i)
1	0.93	$(0, \frac{1}{5})$
1	0.72	$(0, \frac{2}{5})$
1	0.62	$(0, \frac{3}{5})$
1	0.45	$(0, \frac{4}{5})$
0	0.39	$(\frac{1}{3}, \frac{4}{5})$
0	0.32	$(\frac{2}{3}, \frac{4}{5})$
1	0.18	$(\frac{2}{3}, 1)$
0	0.01	$(1, 1)$

那么 $AUC_{C_1} = \frac{1}{2} \sum_{i=1}^7 (x_{i+1} - x_i)(y_i + y_{i+1}) = \frac{13}{15}$ 。

② C_2 :

y	yC_2	(x_i, y_i)
1	0.97	$(0, \frac{1}{5})$
1	0.89	$(0, \frac{2}{5})$
1	0.82	$(0, \frac{3}{5})$
0	0.75	$(\frac{1}{3}, \frac{3}{5})$
0	0.36	$(\frac{2}{3}, \frac{3}{5})$
1	0.34	$(\frac{2}{3}, \frac{4}{5})$
1	0.17	$(\frac{2}{3}, 1)$
0	0.12	$(1, 1)$

那么 $AUC_{C_2} = \frac{1}{2} \sum_{i=1}^7 (x_{i+1} - x_i)(y_i + y_{i+1}) = \frac{11}{15}$.

2)

①当 C_1 设定阈值为0.40时，其混淆矩阵为

真实情况/预测结果	P	N
正例	4	1
反例	0	3

$$P = \frac{TP}{TP+FP} = 1$$

$$R = \frac{TP}{TP+FN} = \frac{4}{5}$$

$$F_1 = \frac{2 \times P \times R}{P+R} = \frac{8}{9}$$

②当 C_2 设定阈值为0.90时，其混淆矩阵为

真实情况/预测结果	P	N
正例	1	4
反例	0	3

$$P = \frac{TP}{TP+FP} = 1$$

$$R = \frac{TP}{TP+FN} = \frac{1}{5}$$

$$F_1 = \frac{2 \times P \times R}{P+R} = \frac{1}{3}$$

2 Linear model

1)

记(2.1)式为 F_w ,

对(2.1)式求导得

$$\begin{aligned}\frac{\partial F_w}{\partial w} &= X^T(Xw - y) + 2\lambda w \\ &= (X^T X + 2\lambda E)w - 2X^T y \quad (2.2)\end{aligned}$$

令(2.2)为0, 由于 X 为列满秩, 故 $X^T X$ 为正定矩阵, 解得通式解如下:

$$w^* = (X^T X + 2\lambda E)^{-1} X^T y \quad (2.2)$$

E 为与 $X^T X$ 同阶单位矩阵.

2)

$\lambda = 1$ 时, (2.2)可写作

$$w^* = (X^T X + 2E)^{-1} X^T y \quad (2.3)$$

由训练集有

$$X = \begin{bmatrix} 2 & 9 & 1 \\ 9 & 3 & 1 \\ 8 & 3 & 1 \\ 8 & 8 & 1 \\ 2 & 1 & 1 \\ 8 & 4 & 1 \\ 4 & 3 & 1 \\ 1 & 8 & 1 \\ 3 & 3 & 1 \\ 5 & 3 & 1 \end{bmatrix}$$

$$y = \begin{bmatrix} 290 \\ 1054 \\ 944 \\ 964 \\ 246 \\ 948 \\ 488 \\ 167 \\ 370 \\ 598 \end{bmatrix}$$

代入(2.3)解得

$$w = \begin{bmatrix} 112.9340 \\ 6.1899 \\ 11.9795 \end{bmatrix}$$

3 Logistic Regression

1) 对式 (3.2) 求二阶导:

$$\frac{\partial^2 \ell(\beta)}{\partial \beta^2} = \sum_{i=1}^m \left(0 + \frac{\partial^2 (\ln(1+e^{\beta^T \hat{x}_i}))}{\partial \beta^2} \right) = \sum_{i=1}^m \frac{\hat{x}_i^2 e^{\beta^T \hat{x}_i}}{(1+e^{\beta^T \hat{x}_i})^2} \geq 0$$

故式 (3.2) 具有凸函数性质。

2)

当 $y_i \in \{1, 2, \dots, K\}$ 时, 有

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_K \end{pmatrix}, \mathbf{W} = (w_1, w_2, \dots, w_K), \mathbf{b} = (b_1, b_2, \dots, b_K)$$

\mathbf{y} 的预测值为 $\mathbf{z} = \mathbf{W}^T \mathbf{x} + \mathbf{b}$,

该模型的对数似然为

$$\begin{aligned} \ell(\mathbf{W}, \mathbf{b}) &= \sum_{i=1}^m \ln p(\mathbf{y}_i | \mathbf{x}_i) \\ &= \sum_{i=1}^m \ln \prod_{j=1}^K p(y_{ij} | \mathbf{x}_i) \\ &= \sum_{i=1}^m \sum_{j=1}^K \ln p(y_{ij} | \mathbf{x}_i) \\ &= \sum_{i=1}^m \sum_{j=1}^K (y_{ij}(\mathbf{w}_j^T \mathbf{x}_i + b_j) - \ln(1 + e^{w_j^T x_i + b_j})) \end{aligned}$$

3)

使用 `sklearn` 工具对给定数据集¹采用OvO,OvR,MvMLR模型进行训练。训练集数据占数据集的 $\frac{7}{10}$,数据集的其余 $\frac{3}{10}$ 数据作训练集。相关代码²

训练结果如下:

模型	训练集正确率(%)	测试集正确率(%)	耗时(s)
OvO	52.8	53.1	0.183
OvR	54.1	52.5	0.058
MvM	53.3	52.5	0.077

各模型的正确率相近。由于OvO模型需要训练 $O(\frac{N(N-1)}{2})$ 个分类器, 耗时较长。OvO模型和MvM模型的耗时相近。