F14081046 系統112 周昱陽.

Problem3.

(c) Support Vector Machine (SVM) Classifier (10%)

Follow the steps on the notebook HW1.ipynb to build a SVM Classifier for MNIST dataset. Here you should use the representative data (after PCA) for training and inference. For more details, please refer to https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html

1) Implement a SVM classifier using the scikit-learn package: sklearn.svm.SVC with L2 regularization parameters C = 1.0, kernel type 'linear'.

2) Evaluate the classification accuracies on the validation set.

3) Try using a different kernel type, change the kernal from 'linear' to 'rbf' (radial basis function) and evaluate the classification accuracy on the validation set. Which one ('linear' or 'rbf') can give you higher accuracy? A: rbf.

4) Fix the kernel type to be 'rbf' and try different sets of regularization parameters $C \in \{0.1, 0.5, 1.0, 5.0, 10.0\}$, and report all the classification accuracies on the validation set. What's the meaning of changing the C here? Which C in your case can give your the best accurarcy?

① we can find better performance on accuracy.

② C = 10 has the best performance.

Problem 4.

1. (5%) Given a trained classifier for 4 object classes (C_1, C_2, C_3 C_4), an input data belongs to C2 generates (0.15, 0.7, 0.1, 0.05) output likelihood, what are the corresponding loss values (L1 loss, L2 loss and cross-entropy loss.) associated with this data?

Y_pred = {0.15, 0.7, 0.1, 0.05}
Y_true = {0, 1, 0, 0}
L1 = sum_{i:1~N}|y_true - y_pred| = |0.15 - 0| + |0.7-1| + |0 - 0.1| + |0-0.05| = ?

$$\text{L1-loss} = \frac{\sum|y_i - \hat{y}_i|}{n} = \frac{0.15 + 0.3 + 0.1 + 0.05}{4} = 0.15 \ \#$$

$$\text{L2-loss} = \frac{\sum(y_i - \hat{y}_i)^2}{n} = \frac{0.15^2 + 0.3^2 + 0.1^2 + 0.05^2}{4} = 0.03125 \ \#$$

$$\text{CrossEntropy loss} = -(y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i))$$

$$= -(1 \cdot \log 0.7 + 0)$$

$$= -0.155 \ \#$$

2、

C_1 : $TP = 68$, $TN = (74+5+7+3+82+3+10+12+72) = 268$.

$FP = 14+12+6 = 32$, $FN = 12+9+11 = 32$.

$Pre = \frac{TP}{TP+FP} = \frac{68}{68+32} = 0.68$. $F_1 = \frac{2TP}{2TP+FP+FN} = \frac{2\cdot 68}{2\cdot 68+32+32} = 0.68$.

$Rec = \frac{TP}{TP+FN} = \frac{68}{68+32} = 0.68$. , $ACC = \frac{TP+TN}{TP+TN+FP+FN} = \frac{336}{400} = 0.84$.

C_2 : $TP_2 = 74$, $TN_2 = 68+9+11+12+82+3+6+12+72 = 275$ $FP_2 = 12+3+10 = 25$. $FN_2 = 14+5+7 = 24$.

$Pre = \frac{TP}{TP+FP} = \frac{74}{74+25} = 0.747$ $F_1 = \frac{2TP}{2TP+FP+FN} = \frac{2\cdot 74}{2\cdot 74+25+24} = 0.751$.

$Rec = \frac{TP}{TP+FN} = \frac{74}{74+24} = 0.755$. $ACC = \frac{TP+TN}{TP+TN+FP+FN} = \frac{349}{400} = 0.8725$.

C_3 : $TP_3 = 82$, $TN_3 = 68+12+11+14+74+7+6+10+72 = 274$, $FP_3 = 9+5+12 = 26$. $FN_3 = 12+3+3 = 18$.

$Pre = \frac{TP}{TP+FP} = \frac{82}{82+26} = 0.76$ $F_1 = \frac{2TP}{2TP+FP+FN} = \frac{2\cdot 82}{2\cdot 82+26+18} = 0.788$

$Rec = \frac{TP}{TP+FN} = \frac{82}{82+18} = 0.82$. $ACC = \frac{TP+TN}{TP+TN+FP+FN} = \frac{356}{400} = 0.89$.

C_4. : $TP_4 = 72$, $TN = 68+12+9+14+74+5+12+3+82 = 279$, $FP_4 = 11+7+3 = 21$ $FN_4 = 6+10+12 = 28$.

$Pre = \frac{TP}{TP+FP} = \frac{72}{72+21} = 0.774$ $F_1 = \frac{2TP}{2TP+FP+FN} = \frac{2\cdot 72}{2\cdot 72+21+28} = 0.746$

$Rec = \frac{TP}{TP+FN} = \frac{72}{72+28} = 0.72$ $ACC = \frac{TP+TN}{TP+TN+FP+FN} = \frac{351}{400} = 0.875$.

overall ave acc = $(0.84+0.8725+0.89+0.875)/4 = 0.869$ #.

F1 score of C4 = 0.746. #.

micro-ave precision = $\frac{68+74+82+72}{68+74+82+72+32+25+26+21} = 0.74$ #.

3. (5%) Multiple Linear Regression: Given the following dataset with one response variable y and two predictor variables x_1 and x_2:

y	X_1	X_2
140	60	22
155	62	25
159	67	24
179	70	20
192	71	15
200	72	14
212	75	14
215	78	11

Please find the linear regression model $y=b_0+b_1x_1+b_2x_2$, i.e., determine the linear regression coefficients, b_0, b_1 and b_2 based on least squares solution.

$$y = b_0 + b_1 x_1 + b_2 x_2 + e$$

$$e = y - b_0 - b_1 x_1 - b_2 x_2$$

$$S_r = \sum (y - b_0 - b_1 x_1 - b_2 x_2)^2$$

$$\begin{cases} \frac{\partial S_r}{\partial b_0} = -2\sum(y-b_0-b_1x_1-b_2x_2) = 0 \\ \frac{\partial S_r}{\partial b_1} = -2\sum(y-b_0-b_1x_1-b_2x_2)x_1 = 0 \\ \frac{\partial S_r}{\partial b_2} = -2\sum(y-b_0-b_1x_1-b_2x_2)x_2 = 0 \end{cases}$$

$$\begin{cases} \sum y = n \cdot b_0 + b_1 \sum x_1 + b_2 \sum x_2 \\ \sum y \cdot x_1 = b_0 \sum x_1 + b_1 \sum x_1^2 + b_2 \sum x_1 x_2 \\ \sum y \cdot x_2 = b_0 \sum x_2 + b_1 \sum x_1 x_2 + b_2 \sum x_2^2 \end{cases}$$

$$\underbrace{\begin{bmatrix} n & \sum x_1 & \sum x_2 \\ \sum x_1 & \sum x_1^2 & \sum x_1 x_2 \\ \sum x_2 & \sum x_1 x_2 & \sum x_2^2 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix}}_{b} = \underbrace{\begin{bmatrix} \sum y \\ \sum y \cdot x_1 \\ \sum y \cdot x_2 \end{bmatrix}}_{C}$$

$n = 8$, $\sum x_1 x_2 = 9859$

$\sum x_1 = 555$, $\sum y = 1452$

$\sum x_2 = 145$, $\sum y \cdot x_1 = 101895$

$\sum x_1^2 = 38767$, $\sum y \cdot x_2 = 25364$

$\sum x_2^2 = 2823$

$$b^* = (A^T A)^{-1} A^T C = \begin{bmatrix} -6.87 \\ 3.15 \\ -1.66 \end{bmatrix}$$

$$y = -6.87 + 3.15 x_1 - 1.66 x_2 \text{ ※}$$

4. (5%) Explain what is Support Vector Machine (SVM) (3%), and when and how do we use nonlinear SVM (Hint: Kernel Trick) (2%)?

① finding a hyperplane that best seperates the data points of different classe's.

② when data points of different classes can't be seperated by a linear hyperplane, we use Kernel Trick map the data into higher-dimensional space, where the data can be seperated.

5. (5%) Explain what are discriminative and generative classifiers?

discriminative : models the decision boundry directly between classes, based on input features, and giving us a posterior probability.

generative : models the joint probability distribution of the input features and output classes, and giving us a likelihood.