Problem 4: Questions and Answers (25%)

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| = ?

Ans:

$$L1 = \frac{\sum_{i=1}^{n}|y_i - \hat{y_i}|^2}{n} = \frac{|0.15-0| + |0.7-1| + |0-0.1| + |0-0.05|}{4} = \frac{0.6}{4} = 0.15$$

$$L2 = \frac{\sum_{i=1}^{n}(y_i - \hat{y_i})^2}{n} = \frac{(0.15-0)^2 + (0.7-1)^2 + (0-0.1)^2 + (0-0.05)^2}{4} = \frac{0.125}{4}$$
$$= 0.03125$$

$$CrossEntropyLoss = \frac{1}{n}\sum_{i=1}^{n} -\big(y_i \log(\hat{y_i}) + (1-y_i)\log(1-\hat{y_i})\big)$$
$$= \frac{-\log(1-0.15) - \log(0.7) - \log(1-0.1) - \log(1-0.05)}{4}$$
$$\approx \frac{0.6758}{4}$$
$$\approx 0.1690$$

在這裡鍵入方程式。

2. (5%) Given the **confusion matrix** of this 4-class classifier

		Predicted Classes			
Actual Classes		C_1	C_2	C_3	C_4
	C_1	68	12	9	11
	C_2	14	74	5	7
	C_3	12	3	82	3
	C_4	6	10	12	72

Please compute the overall average accuracy. Followed by per class precision and recall, F1 score of C4, and the micro-average precision of all 4 classes.

Ans:

$$Accuracy = \frac{68+74+82+72}{68+12+9+11+14+74+5+7+12+3+82+3+6+10+12+72}$$
$$= \frac{296}{400} = 0.74$$

	C_1	C_2	C_3	C_4
TP	68	74	82	72

TN	$74+5+7+3+82+3+10+12+72 = 268$	$68+12+6+9+82+12+11+3+72 = 275$	$68+14+6+12+74+10+11+7+72 = 274$	$68+12+9+14+74+5+12+3+82 = 279$
FP	$14+12+6 = 32$	$12+3+10 = 25$	$9+5+12 = 26$	$11+7+3 = 21$
FN	$12+9+11 = 32$	$14+5+7 = 26$	$12+3+3 = 18$	$6+10+12 = 28$

	C_1	C_2	C_3	C_4
Precision, $\frac{TP}{TP+FP}$	$\frac{68}{68+32} = \frac{68}{100} = 0.68$	$\frac{74}{74+25} = \frac{74}{99} \approx 0.75$	$\frac{82}{82+26} = \frac{82}{108} \approx 0.76$	$\frac{72}{72+21} = \frac{72}{93} \approx 0.77$
Recall, $\frac{TP}{TP+FN}$	$\frac{68}{68+32} = \frac{68}{100} = 0.68$	$\frac{74}{74+26} = \frac{74}{100} = 0.74$	$\frac{82}{82+18} = \frac{82}{100} = 0.82$	$\frac{72}{72+28} = \frac{72}{100} = 0.72$
F1-score, $\frac{2 \cdot Precision \cdot Recall}{Precision + Recall}$	$\frac{2 \times 0.68 \times 0.68}{0.68+0.68} = 0.68$	$\frac{2 \times 0.75 \times 0.74}{0.75+0.74} \approx 0.74$	$\frac{2 \times 0.76 \times 0.82}{0.76+0.82} \approx 0.79$	$\frac{2 \times 0.77 \times 0.72}{0.77+0.72} \approx 0.74$

$$Micro - AveragePrecision = \frac{TP_{C_1} + TP_{C_2} + TP_{C_3} + TP_{C_4}}{TP_{C_1} + TP_{C_2} + TP_{C_3} + TP_{C_4} + FP_{C_1} + FP_{C_2} + FP_{C_3} + FP_{C_4}}$$
$$= \frac{68+74+82+72}{68+74+82+72+32+25+26+21}$$
$$= \frac{296}{400}$$
$$= 0.74$$

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.

$$X = \begin{bmatrix} 1 & 60 & 22 \\ 1 & 62 & 25 \\ 1 & 67 & 24 \\ 1 & 70 & 20 \\ 1 & 71 & 15 \\ 1 & 72 & 14 \\ 1 & 75 & 14 \\ 1 & 78 & 11 \end{bmatrix}, \mathrm{y} = \begin{bmatrix} 140 \\ 155 \\ 159 \\ 179 \\ 192 \\ 200 \\ 212 \\ 215 \end{bmatrix}$$

$$\hat{b} = (X^T X)^{-1} X^T y$$

$$\hat{b} = \left(\begin{bmatrix} 1 & 60 & 22 \\ 1 & 62 & 25 \\ 1 & 67 & 24 \\ 1 & 70 & 20 \\ 1 & 71 & 15 \\ 1 & 72 & 14 \\ 1 & 75 & 14 \\ 1 & 78 & 11 \end{bmatrix}^T \begin{bmatrix} 1 & 60 & 22 \\ 1 & 62 & 25 \\ 1 & 67 & 24 \\ 1 & 70 & 20 \\ 1 & 71 & 15 \\ 1 & 72 & 14 \\ 1 & 75 & 14 \\ 1 & 78 & 11 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 60 & 22 \\ 1 & 62 & 25 \\ 1 & 67 & 24 \\ 1 & 70 & 20 \\ 1 & 71 & 15 \\ 1 & 72 & 14 \\ 1 & 75 & 14 \\ 1 & 78 & 11 \end{bmatrix}^T \begin{bmatrix} 140 \\ 155 \\ 159 \\ 179 \\ 192 \\ 200 \\ 212 \\ 215 \end{bmatrix}$$

$$\hat{b} = \left(\begin{bmatrix} 8 & \sum_i x_{1i} & \sum_i x_{2i} \\ \sum_i x_{1i} & \sum_i x_{1i}^2 & \sum_i x_{1i} x_{2i} \\ \sum_i x_{2i} & \sum_i x_{1i} x_{2i} & \sum_i x_{2i}^2 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 60 & 22 \\ 1 & 62 & 25 \\ 1 & 67 & 24 \\ 1 & 70 & 20 \\ 1 & 71 & 15 \\ 1 & 72 & 14 \\ 1 & 75 & 14 \\ 1 & 78 & 11 \end{bmatrix}^T \begin{bmatrix} 140 \\ 155 \\ 159 \\ 179 \\ 192 \\ 200 \\ 212 \\ 215 \end{bmatrix}$$

$$\hat{b} = \left(\begin{bmatrix} 8 & 555 & 145 \\ 555 & 38767 & 9859 \\ 145 & 9859 & 2823 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 60 & 22 \\ 1 & 62 & 25 \\ 1 & 67 & 24 \\ 1 & 70 & 20 \\ 1 & 71 & 15 \\ 1 & 72 & 14 \\ 1 & 75 & 14 \\ 1 & 78 & 11 \end{bmatrix}^T \begin{bmatrix} 140 \\ 155 \\ 159 \\ 179 \\ 192 \\ 200 \\ 212 \\ 215 \end{bmatrix}$$

$$\because det\left(\begin{bmatrix} 8 & 555 & 145 \\ 555 & 38767 & 9859 \\ 145 & 9859 & 2823 \end{bmatrix} \right)$$

$$= 8 \times 38767 \times 2823 + 555 \times 9859 \times 145 \times 2 - 145 \times 38767 \times 145 - 9859^2 \times 8 - 555^2 \times 2823$$

$$= 90180$$

$$\hat{b} =$$

$$\frac{1}{90180} \begin{bmatrix} 38767(2823) - 9859^2 & 145(9859) - 555(2823) & 555(9859) - 145(38767) \\ 145(9859) - 555(2823) & 8(2823) - 145^2 & 145(555) - 8(9859) \\ 555(9859) - 145(38767) & 145(555) - 8(9859) & 8(38767) - 555^2 \end{bmatrix} \begin{bmatrix} 1 & 60 & 22 \\ 1 & 62 & 25 \\ 1 & 67 & 24 \\ 1 & 70 & 20 \\ 1 & 71 & 15 \\ 1 & 72 & 14 \\ 1 & 75 & 14 \\ 1 & 78 & 11 \end{bmatrix}^T \begin{bmatrix} 140 \\ 155 \\ 159 \\ 179 \\ 192 \\ 200 \\ 212 \\ 215 \end{bmatrix}$$

$$\hat{b}^{[1]} \approx \begin{bmatrix} 135.7214 & -1.5215 & -1.6575 \\ -1.5215 & 0.0173 & 0.0178 \\ -1.6575 & 0.0178 & 0.0234 \end{bmatrix} \begin{bmatrix} 1 & 60 & 22 \\ 1 & 62 & 25 \\ 1 & 67 & 24 \\ 1 & 70 & 20 \\ 1 & 71 & 15 \\ 1 & 72 & 14 \\ 1 & 75 & 14 \\ 1 & 78 & 11 \end{bmatrix}^T \begin{bmatrix} 140 \\ 155 \\ 159 \\ 179 \\ 192 \\ 200 \\ 212 \\ 215 \end{bmatrix}$$

$$\hat{b} = \begin{bmatrix} 7.9664 & -0.0491 & -5.9991 & -3.9336 & 2.8324 & 2.9684 & -1.5961 & -1.1881 \\ -0.0919 & -0.0039 & 0.0648 & 0.0455 & -0.0262 & -0.0267 & 0.0252 & 0.0237 \\ -0.0747 & 0.0311 & 0.0967 & 0.0565 & -0.0427 & -0.0483 & 0.0051 & -0.0117 \end{bmatrix} \begin{bmatrix} 140 \\ 155 \\ 159 \\ 179 \\ 192 \\ 200 \\ 212 \\ 215 \end{bmatrix}$$

$$\therefore \hat{b} \approx \begin{bmatrix} -6.60 \\ 5.04 \\ 0.56 \end{bmatrix}$$

$$\therefore, \widehat{b_0} = -.6.60, \widehat{b_1} = 5.04, \widehat{b_2} = 0.56$$

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

Ans:

a) Support vector machine is a model that finding a separating hyperplane so that the distances of support vectors of the classes-to-separate to the hyperplane, i.e., margin, satisfy the requirement.
b) Nonlinear SVM is used when the input space can be mapped to some "implicitly" higher-dimensional feature space via nonlinear mapping, making the training set is more separable. To use the nonlinear SVM, we can define a kernel function which is nonlinear and corresponds to the dot product of two feature vectors in some expanded feature space.

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

Discriminative classifiers are the models that requires all the training examples of different classes to be jointly used to build the model up. General classifiers are the models that have to be trained independently on only the examples of the same label.

---

[1] Because of the approximation, the final b is not accurate, the more accurate answer is [-6.87,3.15,-1.66].