Problem set 3 Solution
Normalisation & Classifier evaluation

Excercise 1

Assume that the height of a student takes the following values $[170, 160, 155, 165]$.

1. Use $[0, 1]$-scaling to transform the four points.
2. Use Gaussian normalisation to transform the four data points.

Solution

1. For $[0, 1]$-scaling we map x to $\frac{x-\min}{\max-\min}$. Since $\max = 170, \min = 155$, and $\max - \min = 15$, the scaled vector is $[1, 1/3, 0, 2/3]$.
2. To perform Gaussian normalisation for a vector $(x_1, x_2, \ldots, x_n)$, we map each x_i to $\frac{x_i-\mu}{\sigma}$, where $\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$ is the sample mean and $\sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\mu)^2}$ is the sample standard deviation. Applying this to our vector $[170, 160, 155, 165]$, we find $\mu = 162.5$ and $\sigma \approx 6.4550$.

Excercise 2

A binary classifier was evaluated using a set of 1000 test examples in which 50% of all examples are negative. It was found that the classifier has 0.6 recall and 0.7 accuracy. Write the confusion matrix.

Solution

Let N and P denote the number of negative and positive samples, respectively. Then we know that $N = 500$ and $P = 1000 - 500 = 500$.

We also know that

a. Recall $= \frac{TP}{TP+FN} = \frac{TP}{P} = \frac{TP}{500} = 0.6$, which implies that $TP = 300$.

b. Accuracy $= \frac{TP+TN}{1000} = 0.7$, which implies that $TN = 400$.

c. Since $FN = P - TP$ and $FP = N - TN$, we obtain the following confusion matrix

	Actual Positive	Actual Negative
Predicted Positive	300	100
Predicted Negative	200	400

Excercise 3

Given the confusion matrix for a 3-class classifier

$$\begin{pmatrix} & \text{Car} & \text{Train} & \text{Cycle} \\ \text{Car} & 8 & 3 & 6 \\ \text{Train} & 2 & 4 & 2 \\ \text{Cycle} & 2 & 4 & 12 \end{pmatrix}$$

1. Calculate Precision, Recall, and F-score for each of the 3 classes
2. Calculate Macro F-score

Solution

- For class Car: Precision = 8/17, Recall = 2/3, F-Score = 16/29
- For class Train: Precision = 1/2, Recall = 4/11, F-Score = 8/19
- For class Cycle: Precision = 2/3, Recall = 3/5, F-Score = 12/19

Macro F-score is equal to $\frac{1}{3}(16/29 + 8/19 + 12/19) \approx 0.53$.