Exercise sheet: End-to-end ML project

The following exercises have different levels of difficulty indicated by (*), (**), (***). An exercise with (*) is a simple exercise requiring less time to solve compared to an exercise with (***), which is a more complex exercise.

1. (*) You have built an ML classifier that detects whether a tissue appearing in an image is cancerous or not. Consider the cancerous class as the positive class. The following confusion matrix shows the predicted results obtained in the validation set

	cancerous (predicted)	healthy (predicted)
cancerous (actual)	30	5
healthy (actual)	15	100

   Compute the precision, recall and accuracy of your ML classifier.

2. (*) Table 1 below shows the scores achieved by a group of students on an exam. Using this data, perform the following tasks on the Score feature
   (a) A normalisation in the range $[0, 1]$.
   (b) A normalisation in the range $[-1, 1]$.
   (c) A standardisation of the data.

ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Score	42	47	59	27	84	49	72	43	73	59	58	82	50	79	89	75	70	59	67	35

Table 1: Students' score

3. (*) We designed a model for predicting the number of bike rentals (y) from two attributes, temperature (x_1) and humidity (x_2),

$$y = 500 \times x_1 + 300 \times x_2.$$

   The model was trained with normalised data with values $\min x_1 = -10$ and $\max x_1 = 39$ for x_1, and values $\min x_2 = 20$ and $\max x_2 = 100$. At test time, the model is used to predict the bike rentals for a vector $\mathbf{x}_* = [25,\ 70]^\top$. What is the value of the prediction y?

4. (*) A simple criterion to remove outliers from a dataset is to compute the mean, μ, and the standard deviation, σ, of the variable of interest and consider values outside the range $(\mu - 3\sigma, \mu + 3\sigma)$ as outliers. Applying this criterion to the Scores in Exercise 2, which ones of them can be considered as outliers?

5. (**) Suppose the joint pmf of the two RVs X and Y is given as

$$P(X = x_i, Y = y_j) = \begin{cases} \frac{1}{3}, & \text{for } (x_1 = 0, y_1 = 1), (x_2 = 1, y_2 = 0), (x_3 = 2, y_1 = 1) \\ 0 & \text{otherwise,} \end{cases}$$

(a) Are X and Y independent?

(b) Are X and Y uncorrelated?

6. (**) Two RVs X and Y are uncorrelated if $\sigma_{X,Y} = 0$. Since $\sigma_{X,Y} = E\{XY\} - E\{X\}E\{Y\}$, the two RVs are uncorrelated if $E\{XY\} = E\{X\}E\{Y\}$. Show that if the RVs are independent, then they are also uncorrelated.
$\big[$HINT: the expected value $E\{XY\}$ is defined as

$$E\{XY\} = \sum_{\forall x_i} \sum_{\forall y_j} x_i y_j P(x_i, y_j),$$

where $P(x_i, y_j)$ is the joint pmf for the discrete RVs X and Y. A similar definition can be written if X and Y are continuous RVs, replacing the sums for integrals. $\big]$

7. (***) Let $Y = aX + b$, where Y and X are RVs and a and b are constants.

(a) Find the covariance of X and Y.

(b) Find the correlation coefficient of X and Y.