Machine Learning Exercises 2

Supervised vs Unsupervised problems

Exercise 1. For each of the following problems, determine whether the problem is supervised or unsupervised.

(a) Given detailed phone usage from many people, find interesting groups of people with similar behaviour.
(b) Given detailed phone usage of many users along with their historic churn, predict if people are going to change contracts again.
(c) Given expression measurements of 1000s of genes for 100s of patients along with a binary variable indicating presence or absence of a specific cancer, predict if the cancer is present for a new patient.
(d) Given expression measurements of 1000s of genes for 100s of patients, find groups of functionally similar genes.

Classification vs Regression

Exercise 2. For data with each of the following outcome variables, determine whether the problem is suitable for classification or regression:

(a) Presence or absence of cancer.
(b) Favourite fruits
(c) Annual income in kroner
(d) Income bracket

Linear regression models

Imagine we have a dataset with two features: x_1 and x_2 that are numerical (real-valued) variables

Consider the following models,

(a) $Y = \beta_0 + \beta_1 x_1 + \varepsilon$

(b) $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$

(c) $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 (x_1 x_2) + \varepsilon$

(d) $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 (x_1 x_2) + \beta_4 x_1^2 + \beta_5 x_2^2 + \varepsilon$

Exercise 3. Explain how a unit change in x_1 would affect y (leaving x_2 unchanged) in each model (a–d)

Exercise 4. Make a hand-drawn sketch of the functional relationship between Y and the two features in each of the models (a–d)

Exercise 5. A linear regression model can be formulated also in matrix notation as

$$Y = X\beta + \varepsilon.$$

Explain what X would be for each of models (a–d).

Exercise 6. Introduce now a third, categorical, feature C with two levels yes/no.

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 C_{\text{yes}} + \varepsilon$$

Exercise 7. Sketch and explain the change to the relationship between Y and x_1 and x_2 if you introduce interactions between C and x_1 and between C and x_2 in each of the four models (a–d).