I2ML – Test exam – WS2021/22

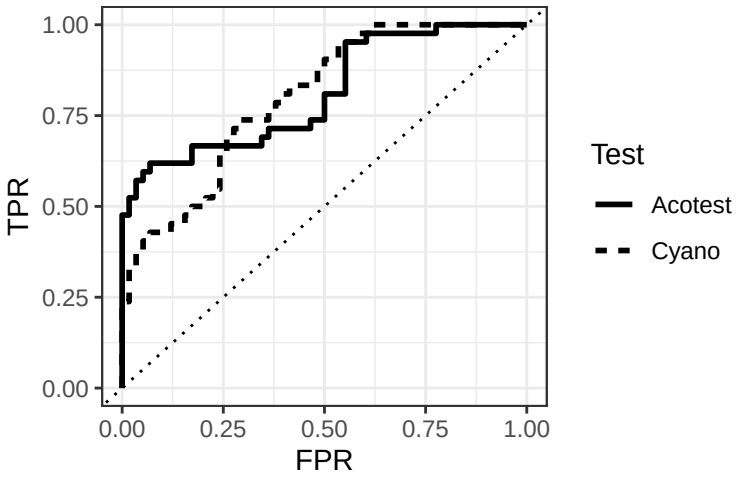
Exercise 1: Polynomial regression

State the hypothesis space for a univariate polynomial regression model with

- degree $d = 4$ and
- first-order interactions between all terms of even order.

Exercise 2: ROC

A small island nation is challenged by a newly emerged respiratory virus. Fortunately, there have been no recorded cases in the country so far. Therefore, the island's government wants to implement a test-on-arrival screening program that ensures that as many virus-positive travelers as possible are identified at the island's airport by mandatory rapid antigen tests and can be subsequently quarantined. For this purpose, you are tasked with selecting the test that is most suited for these circumstances. The manufacturer has provided you with cross-validated ROC curves for the two tests he has developed.

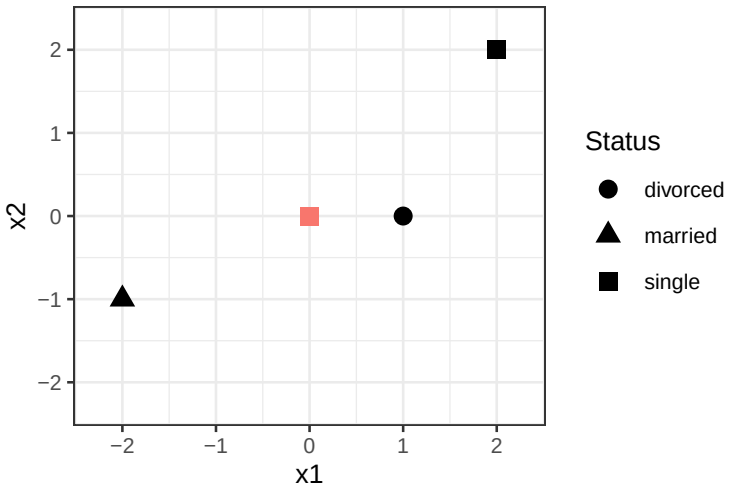


a) Which of the two tests would you recommend given the government's objective? Explain your choice.

b) The country's health minister is skeptical of a testing regime that wrongly classifies travelers as virus-positive and quarantines them for two weeks. She demands a testing regime to be chosen that ensures no one is wrongly quarantined. Which test seems best suited for her demand? Under such a regime, what is the maximum fraction of virus-positive travelers that can be filtered out by the screening program?

Exercise 3: k-NN

You are given data with observations $\mathbf{x}^{(i)} = (x_1,\ x_2,\ x_{\text{status}})^{(i)} \in \mathcal{X}$ with $x_{\text{status}} \in \{\text{single}, \text{married}, \text{divorced}\}$ as plotted below. Find the k-neighborhood of the red point $\mathbf{x}^* = (0,\ 0,\ \textit{single})$ for $k = 1$ using a distance measure from the lecture that is suited to the data format.



Exercise 4: LDA

Derive the LDA decision boundary for a binary problem in two dimensions. For this, employ the alternative description of the LDA decision rule we (implicitly) used to demonstrate LDA's linearity:

$$\delta_k(\mathbf{x}) = \log \pi_k - \frac{1}{2}\boldsymbol{\mu_k}^\top \Sigma^{-1} \boldsymbol{\mu_k} + \mathbf{x}^\top \Sigma^{-1} \boldsymbol{\mu_k}, \quad k \in \{1, 2\}.$$

Explicitly state the equation for the boundary between the two classes.