

I2ML – Test exam – WS2022/23

Solution 1: Polynomial regression

(a) With $\boldsymbol{\theta} := (\theta_0, \theta_1, \theta_2, \theta_3, \theta_4)^\top$, we have

$$\mathcal{H} = \{f(\mathbf{x} \mid \boldsymbol{\theta}) = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4 \mid \boldsymbol{\theta} \in \mathbb{R}^5\}$$

(b) With $\boldsymbol{\theta} := (\theta_0, \theta_1, \theta_2, \theta_3, \theta_4, \theta_5)^\top$, we have

$$\begin{aligned}\mathcal{H} &= \{f(\mathbf{x} \mid \boldsymbol{\theta}) = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4 + \theta_5 x^2 x^4 \mid \boldsymbol{\theta} \in \mathbb{R}^6\} \\ &= \{f(\mathbf{x} \mid \boldsymbol{\theta}) = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4 + \theta_5 x^6 \mid \boldsymbol{\theta} \in \mathbb{R}^6\}\end{aligned}$$

Solution 2: ROC

- a) The objective of identifying as many virus-positive travelers as possible is best served with the Cyano test. Since the *TPR* has to be 1.00, Cyano allows for the threshold to be chosen to achieve a lower *FPR* of roughly 0.625, whereas Acotest only achieves an *FPR* of about 0.78 for the same *TPR*.
- b) Acotest is best suited for this demand. Since the *FPR* has to be 0.00, it allows for the threshold to be chosen to achieve a higher *TPR* than Cyano. In this case, a maximum of roughly 50 percent of all virus-positive travelers can be expected to be filtered out by the testing regime.

Solution 3: *k*-NN

As x_{status} is a categorical feature, the gower distance is suited as a distance measure:

$$d_{gower}(\mathbf{x}, \tilde{\mathbf{x}}) = \frac{\sum_{j=1}^p \delta_{x_j, \tilde{x}_j} \cdot d_{gower}(x_j, \tilde{x}_j)}{\sum_{j=1}^p \delta_{x_j, \tilde{x}_j}}$$

Gower distance for \mathbf{x}^* and $\mathbf{x}^{(1)} = (-2, -1, married)$:

$$d_{gower}(\mathbf{x}^*, \mathbf{x}^{(1)}) = \frac{1 \cdot \frac{|-2-0|}{|-2-2|} + 1 \cdot \frac{|-1-0|}{|-1-2|} + 1 \cdot 1}{1 + 1 + 1} = \frac{\frac{2}{4} + \frac{1}{3} + 1}{3} = 0.611$$

Gower distance for \mathbf{x}^* and $\mathbf{x}^{(2)} = (1, 0, divorced)$:

$$d_{gower}(\mathbf{x}^*, \mathbf{x}^{(2)}) = \frac{1 \cdot \frac{|1-0|}{|-2-2|} + 1 \cdot \frac{|0-0|}{|-1-2|} + 1 \cdot 1}{1 + 1 + 1} = \frac{\frac{1}{4} + \frac{0}{3} + 1}{3} = 0.417$$

Gower distance for \mathbf{x}^* and $\mathbf{x}^{(3)} = (2, 2, single)$:

$$d_{gower}(\mathbf{x}^*, \mathbf{x}^{(3)}) = \frac{1 \cdot \frac{|2-0|}{|-2-2|} + 1 \cdot \frac{|2-0|}{|-1-2|} + 1 \cdot 0}{1 + 1 + 1} = \frac{\frac{2}{4} + \frac{2}{3} + 0}{3} = 0.389$$

Therefore, the 1-neighborhood $N_1(\mathbf{x}^*)$ of the red point \mathbf{x}^* is the point $\mathbf{x}^{(3)}$, which is the observation with the lowest gower distance.

Solution 4: LDA

In order to arrive at the equation for the decision boundary, we first need to understand that, on the boundary of classes 1 and 2, both discriminant functions $\delta_1(\mathbf{x})$ and $\delta_2(\mathbf{x})$ will be exactly equal. Therefore, we compute the equation as follows:

$$\begin{aligned}\delta_1(\mathbf{x}) &= \delta_2(\mathbf{x}) \\ \Leftrightarrow \log \pi_1 - \frac{1}{2} \boldsymbol{\mu}_1^\top \Sigma^{-1} \boldsymbol{\mu}_1 + \mathbf{x}^\top \Sigma^{-1} \boldsymbol{\mu}_1 &= \log \pi_2 - \frac{1}{2} \boldsymbol{\mu}_2^\top \Sigma^{-1} \boldsymbol{\mu}_2 + \mathbf{x}^\top \Sigma^{-1} \boldsymbol{\mu}_2 \\ \Leftrightarrow \mathbf{x}^\top \Sigma^{-1} \boldsymbol{\mu}_1 - \mathbf{x}^\top \Sigma^{-1} \boldsymbol{\mu}_2 &= \log \frac{\pi_2}{\pi_1} + \frac{1}{2} (\boldsymbol{\mu}_1^\top \Sigma^{-1} \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2^\top \Sigma^{-1} \boldsymbol{\mu}_2) \\ \Leftrightarrow \mathbf{x}^\top \underbrace{(\Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2))}_{=: \boldsymbol{\nu} \in \mathbb{R}^{2 \times 1}} &= \log \frac{\pi_2}{\pi_1} + \frac{1}{2} (\boldsymbol{\mu}_1^\top \Sigma^{-1} \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2^\top \Sigma^{-1} \boldsymbol{\mu}_2) \\ \Leftrightarrow \mathbf{x}^\top \boldsymbol{\nu} &= \underbrace{\log \frac{\pi_2}{\pi_1} + \frac{1}{2} (\boldsymbol{\mu}_1^\top \Sigma^{-1} \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2^\top \Sigma^{-1} \boldsymbol{\mu}_2)}_{=: a \in \mathbb{R}}.\end{aligned}$$

The right hand side might look somewhat complicated but simply evaluates to a scalar and we obtain the hyperplane equation $\mathbf{x}^\top \boldsymbol{\nu} = a$, in this case defining a line in \mathbb{R}^2 .

Again, we see that LDA is indeed a linear classifier.