Concept Learning & Evaluation

Exercise 1 : Rule-Based Learning in 2D

Consider the problem of rule-based learning in the following, rather different, feature space: The set of possible examples is given by all points of the x-y plane with integer coordinates from the interval $[1, 10]$. The hypothesis space is given by the set of all rectangles. A rectangle is defined by the points (x_1, y_1) and (x_2, y_2) (bottom left and upper right corner). Hypotheses are written as $\theta = (x_1, y_1, x_2, y_2)$, and assign a point (x, y) to the value 1, if $x_1 \leq x \leq x_2$ and $y_1 \leq y \leq y_2$ hold, with arbitrary, but fixed integer values for x_1, y_1, x_2, y_2 from the interval $[1, 10]$.

Hint: The maximally specific hypothesis h_{s_0} corresponds to a "zero-sized" rectangle that doesn't contain any points with integer coordinates; you may use the symbol $(\perp, \perp, \perp, \perp)$.

(a) For the setting described above, formulate the most general hypothesis h_{g_0}.

(b) Clarify for yourself how the "more-general" relation $\geq_g$ works in this setting, and check all that apply:

- ☐ $(1, 2, 3, 4) \geq_g (1, 1, 4, 4)$
- ☐ $(2, 3, 6, 7) \geq_g (3, 4, 5, 7)$
- ☐ $(1, 1, 2, 8) \geq_g (1, 1, 3, 3)$
- ☐ $(3, 3, 9, 9) \geq_g (1, 1, 1, 1)$

(c) Given a hypothesis $h : \theta = (2, 3, 5, 7)$, and an example $\mathbf{x} = (2, 7)$ with $c = 0$, determine two hypotheses h_1 and h_2 such that both are minimal specializations of h, and both are consistent with $(\mathbf{x}, c)$.

Hint: for the correct answers h_i, there must not exist any other hypothesis h' consistent with $(\mathbf{x}, c)$ where $h \geq_g h'$ and $h' \geq_g h_i$.

Exercise 2 : Precision and Recall

In which of the following classification tasks do we aim for high precision, in which for high recall? Why?

(a) Explosive detection using an airport x-ray machine.

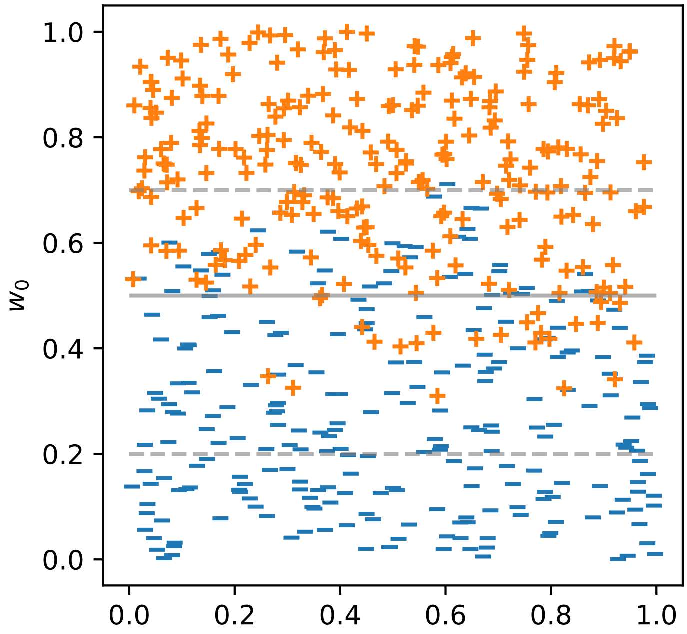
(b) Youtube video recommendations (classifying videos as relevant).

(c) Choosing a good seat on a half-full train.

(d) Spell checking (spelling error detection).

Exercise 3 : Receiver Operating Characteristic (ROC)

Consider the following binary classification scenario, in which each point corresponds to an example $(\mathbf{x}, c)$ with classes $c \in \{0, 1\}$, represented by $-$ and $+$:



We want to examine different linear classifiers $y(\mathbf{x}) = w_0$ (horizontal lines) by calculating the effect of the choice of w_0 on the following two performance metrics:

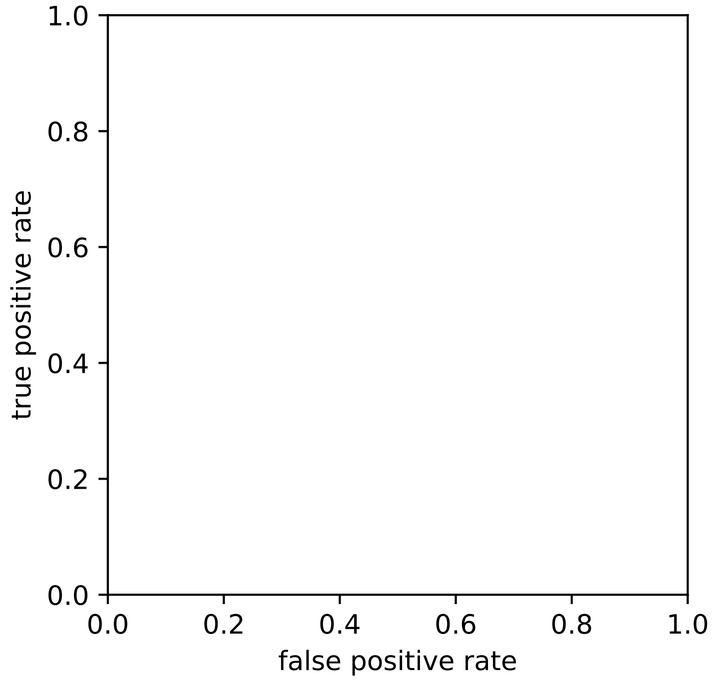
- the *false positive rate*, defined as

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}} = \frac{\text{FP}}{\text{N}} = \frac{|\{(\mathbf{x}, c) \in D : y(\mathbf{x}) = 1 \wedge c = 0\}|}{|\{(\mathbf{x}, c) \in D : c = 0\}|},$$

- and the *true positive rate* (also known as *sensitivity* or *recall*), defined as

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} = \frac{\text{TP}}{\text{P}} = \frac{|\{(\mathbf{x}, c) \in D : y(\mathbf{x}) = 1 \wedge c = 1\}|}{|\{(\mathbf{x}, c) \in D : c = 1\}|}.$$

(a) The *receiver operating characteristic* (ROC) curve describes the relationship between the *true positive* and the *false positive* rate of a binary classifier at different decision thresholds. Varying the value of w_0 in the interval $[0, 1]$, sketch the ROC curve for the classifier above.



(b) What would the ROC curve of a slightly worse (but better than random guess) classifier look like?

(c) What does the ROC curve of the optimal classifier look like?

(d) What does the ROC curve of the worst possible classifier look like? What went wrong? How could this classifier be rectified?

(e) What does the ROC curve of a random classifier look like that uses w_0 as its decision probability?

(f) How does this relate to forming a classifier from a regression model? Use the terms of bias and threshold.