Please carefully read and follow the general instructions regarding exercises. Failing to meet the requirements might lead to penalties. https://elearn.uef.fi/mod/page/view.php?id=293750

If you suspect that something is wrong with some exercise question, please contact the lecturer.

If you face persistent issues while working on an exercise, do ask for help, e.g. during a course meeting or by contacting the lecturer via email.

Consider the dataset consisting of twelve training instances (ids 1–12, on the left) and seven test instances (ids 13–19, on the right), with four variables v_1–v_4 and a class label y that can take one of three values, shown in Figure 1.

id	v_1	v_2	v_3	v_4	y
(1)	1	4	−1	0	●
(2)	3	6	−2	0	●
(3)	7	5	−6	0	●
(4)	2	5	1	0	●
(5)	0	4	6	0	●
(6)	4	6	2	0	●
(7)	6	2	−1	0	●
(8)	8	3	−6	0	●
(9)	7	1	1	1	●
(10)	3	2	6	1	●
(11)	5	1	2	1	●
(12)	1	3	2	1	●

id	v_1	v_2	v_3	v_4	y
(13)	7	4	−6	0	●
(14)	3	3	1	0	●
(15)	4	5	−2	0	●
(16)	6	5	6	1	●
(17)	0	3	−1	0	●
(18)	9	2	1	1	●
(19)	1	1	6	0	●

Figure 1: Dataset consisting of twelve training instances (left) and seven test instances (right)

Problem 1 (Evaluation measures).

a) Looking back at the predictions for the test instances obtained with the decision tree and naive Bayes classifiers when considering three distinct classes, write down the contingency matrix and compute the corresponding accuracy for each method.

b) Looking back at the predictions for the test instances obtained with the decision tree, naive Bayes and SVM classifiers when considering the binary problem (red vs. combined blue or yellow), write down the confusion matrix and compute the accuracy, recall and precision for each method.

c) Comment on the obtained values.

Problem 2 (ROC curves).

a) Looking back at the predictions for the test instances obtained with the SVM and naive Bayes classifiers, compute the corresponding receiver operating characteristic (ROC) curves. As scores for transformed test instance $\boldsymbol{v}'$, use respectively the value of $\boldsymbol{w} \cdot \boldsymbol{v}' + b$ where $\boldsymbol{w}$ and b are the weights and bias of the SVM's separating hyperplane, and the Bayesian posterior probability of belonging to the red class $\mathrm{P}(● \mid \boldsymbol{v}')$.
Draw the curves and report the area under the curve (AUC).