Machine learning for data science I

5 June 2024

Surname, name (all caps) _____

Student ID: _____

This is a closed book exam.

Write clearly and justify your answers.

Time limit: 105 min.

Question:	1	2	3	4	5	Total
Points:	20	20	20	20	20	100
Score:						

1. Explain your answers to the following True/False questions about boosting.

[2] (a) AdaBoost is an ensamble method.

[2] (b) AdaBoost with decision stumps (decision trees with a single split) produces a linear classifier.

[2] (c) Boosting reduces the variance while bagging reduces the bias.

[2] (d) AdaBoost can boost any classifier, not just decision trees.

[2] (e) The error rate of an AdaBoost-ed model never increases from one round to the next.

[2] (f) AdaBoost accounts for outliers by lowering the weights of training points that are repeatedly misclassified.

[2] (g) Cross validation can be used to select the number of iterations in boosting to help reduce overfitting.

[2] (h) The form of the decision boundary of a boosted model is the same as that of a weak learner used in the boosting procedure but with different parameters.

[2] (i) The AdaBoost algorithm is guaranteed to assign the highest weight in the final ensemble model to the weak learner that performs the best on the training set.

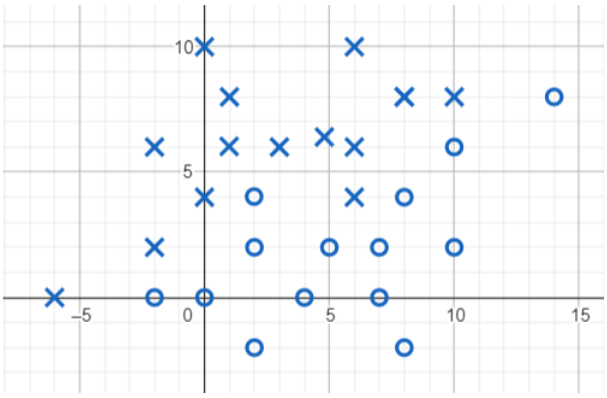
[2] (j) Gradient boosting optimizes an arbitrary differentiable loss function with a sequence of decision trees.

2. Answer the following questions about Logistic Regression.

[6] (a) Describe the Logistic Regression model.

[4] (b) Determine/estimate the decision boundary of logistic regression for the data set illustrated below - write down the equation of the boundary. Crosses are positive instances and circles negative.

[10] (c) Determine/estimate the coefficients of the logistic regression for the data set below - write down the actual fitted model. The probability of the positive class for instance (-2,0) is 15%. Describe your solving process. You don't have to simplify the expression to numeric values.



3. RBF kernel.

[5] (a) Explain which functions can be used as (valid) kernels?

[5] (b) Show that adding a positive constant to some valid kernel function results in a new valid kernel function. Constructing an explicit transformation.

[10] (c) Prove that the RBF kernel $k(x, x') = \exp(-\frac{1}{2\sigma^2}||x - x'||^2)$ is a valid kernel function.

4. You are building a classification model that will be able to correctly classify two types of fish that you catch (there are other types of fish that you can classify perfectly). We will denote with p_1 and p_2 the probabilities of catching each type of fish. The only feature that you observe is the length of the fish. In standardized units the distribution of length is $N(0, \sigma^2)$ for the first and $N(1, \sigma^2)$ for the second type of fish.

   The classification model will be used to sort the fish into different containers. Misclassification cost $\lambda_{a,b} \geq 0$ is the cost of classifying a fish as type a when in fact it's of type b. Correct classifications have a cost $0 = \lambda_{1,1} = \lambda_{2,2}$.

[3] (a) What kind of assumptions did we make with the chosen distributions of length (parameters 0, 1 and σ^2) - how general is such set up?

[7] (b) Write the expression for the risk (expected misclassification cost).

[10] (c) Compute the threshold τ which minimizes the risk.

5. Consider multi-dimensional scaling (MDS).

[4] (a) define the inputs and the outputs of this method,

[4] (b) mathematically formulate the optimization problem,

[4] (c) propose the solution of the optimization problem using the gradient descent approach,

[4] (d) derive the gradients needed,

[4] (e) comment on potential deficiencies when comparing this method to embedding with t-SNE.