(20p) Multiple choice questions (new questions):

* boosting algorithms are easy to parallelize
* the entropy of a dataset is lowest when the classes are balanced
* PRISM algorithm covers only one class at a time

(2p) Perform 1R boosting on the following dataset. Circle the observation with the biggest weight after the first iteration:

X X

O

X X

(2p) Draw the SVM margin of the following dataset. What is the slop of the margin?

O

O

X

X

(4p) Describe Bayesian optimization for hyperparameter tuning

(2p) Training/Validation error curves (error vs model complexity graph). Identify which curve is most likely the training respectively the validation error. Indicate the regions where the model is underfitting respectively overfitting.

(4p) Explain gradient descent for linear regression (compare with other methods; advantages and disadvantages)

(2p) Circle the models that achieve 0 training error on the following dataset

X O

O X

Logistic regression

SVM (polynomial kernel)

Decision tree

1-NN

(8p) X^2 = (|N01 – N10|-1)^2/(N01 + N10)

(3p) What test is this? What is it use for? Explain another significant test.

(5p) Two experiment confusion tables are provied

* + - Calculate accuracy
    - Perform test above? Which model is better? (chi-square table is provided)

(16p) low-medium-high risk dataset with features income, collateral, credit history and debt. Predict the unlabelled observations with: (a) 1R (8p); (b) Bayesian Network (8p)