# Georgia State University Computer Science

# CSC 4850 / 6850– Practice Midterm Machine Learning

### Instructor: Prof. Dong Hye Ye

### 2023/2/23

**Name**: _____

**Student Number**: _____

This exam contains 7 pages (including this cover page) and 4 questions. Total of points is 100. Good luck!

### Distribution of Marks

| Question | Points | Score |
|:---:|:---:|:---:|
| 1 | 30 | |
| 2 | 20 | |
| 3 | 20 | |
| 4 | 30 | |
| Total: | 100 | |

# 1. **Short Questions [30 points]**

The following short questions should be answered with at most two sentences, and/or a pic- ture. For the (true/false) questions, answer true or false. If you answer true, provide a short justification, if false explain why or provide a small counterexample.

(a) (3 points) Discuss whether MAP estimates are less prone to overfitting than MLE.
True

(b) (3 points) A classifier trained on less training data is less likely to overfit.
False, it is more likely to overfit

(c) (3 points) For an SVM, if we remove one of the support vectors from the training set, does the size of the maximum margin decrease, stay the same, or increase for that data set?
Hyperplane mainly based on support vector so if removed it will either increase or stay the same

(d) (3 points) There is a priori good choice of n for n-fold cross-validation.
False, because n depends on the dataset being used

(e) (3 points) In AdaBoost weights of the misclassified examples go up by the same multiplicative factor.
False, weights are updated based on how each classifier performs

(f) (3 points) AdaBoost will eventually give zero training error regardless of the type of weak classifier it uses, provided enough iterations are performed.
False, it tends to decrease towards zero but not actually get to zero.

(g) (3 points) Consider a point that is correctly classified and distant from the decision boundary. Why would SVM's decision boundary be unaffected by this point, but the one learned by logistic regression be affected?
SVM uses hinge loss and logistic regression uses log loss which does not take into account the distant

(h) (3 points) Why does the kernel trick allow us to solve SVMs with high dimensional feature spaces, without significantly increasing the running time?
***

(i) (3 points) "My algorithm is better than yours. Look at the training error rates!"
False, training error is not representative of true data, could overfit.

(j) (3 points) "My algorithm is better than yours. Look at the test error rates! (Footnote: reported results for $\lambda = 1.789489345672120002$.)"

**False, with just this info it's possible there was no hyperparameter tuning.

CSC 4850 /
6850

Practice
Midterm

2023/2
/23
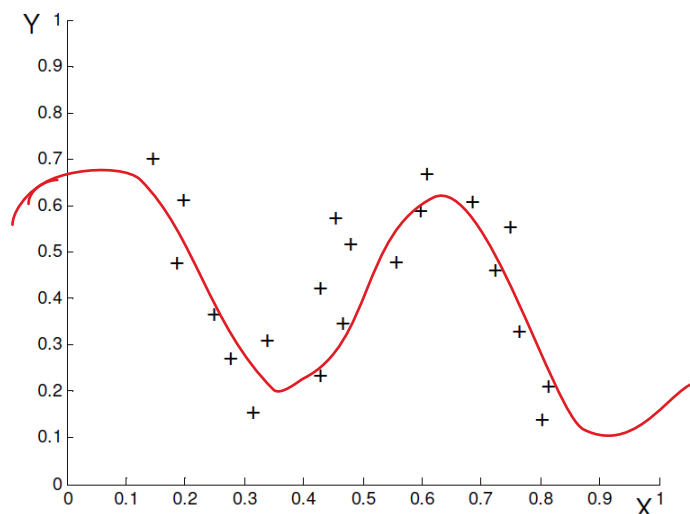
## 2. Regression [20 points]

For each of the following questions, you are given the same data set. Your task is to fit a smooth function to this data set using several regression techniques. Please answer all questions qualitatively, drawing the functions in the respective figures.

(a) (6 points) Show the least squares fit of a linear regression model $Y = aX + b$.
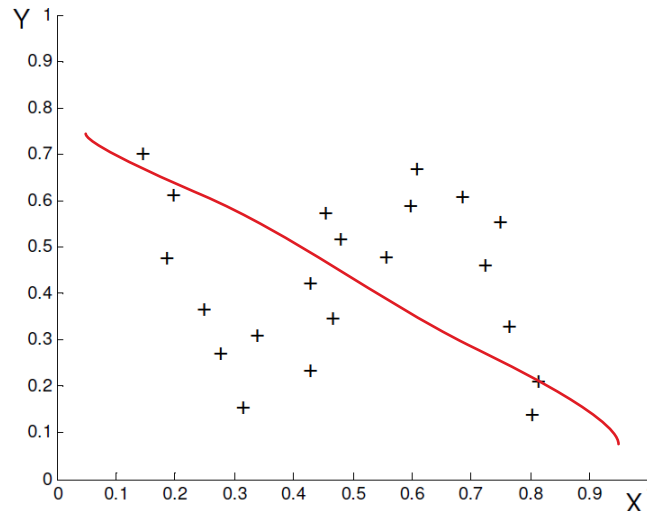


line through middle

(b) (7 points) Show the fit using Kernel regression with Gaussian kernel and an appropriately chosen bandwidth. **Follows Graph and Continues curves at ends**

CSC 4850 /
6850

Practice
Midterm

2023/2
/23

(c) (7 points) Show the fit using Kernel local linear regression for an appropriately chosen bandwidth. **Follows Graph and goes straight at ends**
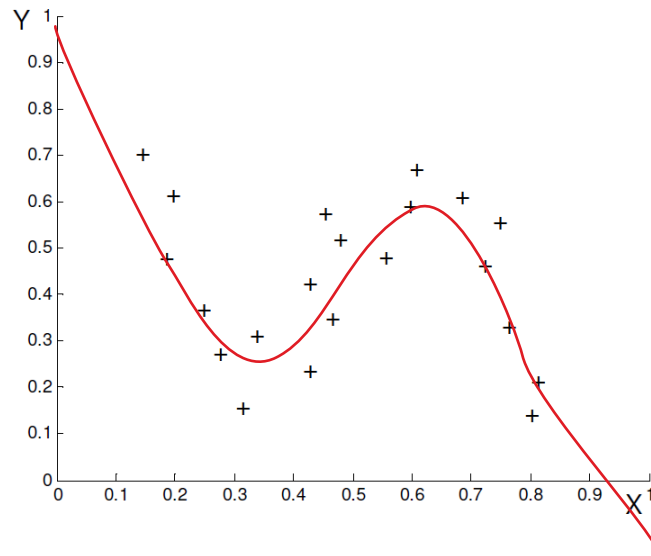
### 3. *k*-nearest neighbor [20 points]

In the following questions you will consider a *k*-nearest neighbor classifier using Euclidean distance metric on a binary classification task. We assign the class of the test point to be the class of the majority of the *k* nearest neighbors. Note that a point can be its own neighbor.
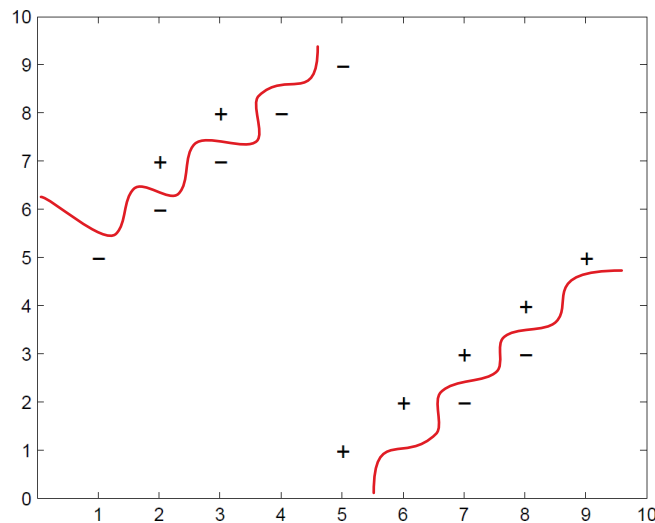


(a) (10 points) Why might using too large values k be bad in this dataset? Why might too small values of k also be bad?
Too large: too high would create a situation where you always have a wrong classification (over-regularized)
Too small: over-fitting

(b) (10 points) In Figure, sketch the 1-nearest neighbor decision boundary for this dataset.

CSC 4850 /
6850

Practice
Midterm

2023/2
/23

## 4. **SVMs and the slack penalty $C$ [30 points]**

The goal of this problem is to correctly classify test data points, given a training data set. You have been warned, however, that the training data comes from sensors which can be error-prone, so you should avoid trusting any specific point too much. For this problem, assume that we are training an SVM with a quadratic kernel– that is, our kernel function is a polynomial kernel of degree 2. You are given the data set presented in Figure. The slack penalty $C$ will determine the location of the separating hyperplane. Please answer the following questions qualitatively. Give a one sentence answer/justification for each and draw your solution in the appropriate part of the Figure at the end of the problem.

(a) (7 points) Where would the decision boundary be for very large values of $C$ (i.e., $C \rightarrow \infty$)? (remember that we are using an SVM with a quadratic kernel.) Draw on the figure below. Justify your answer.
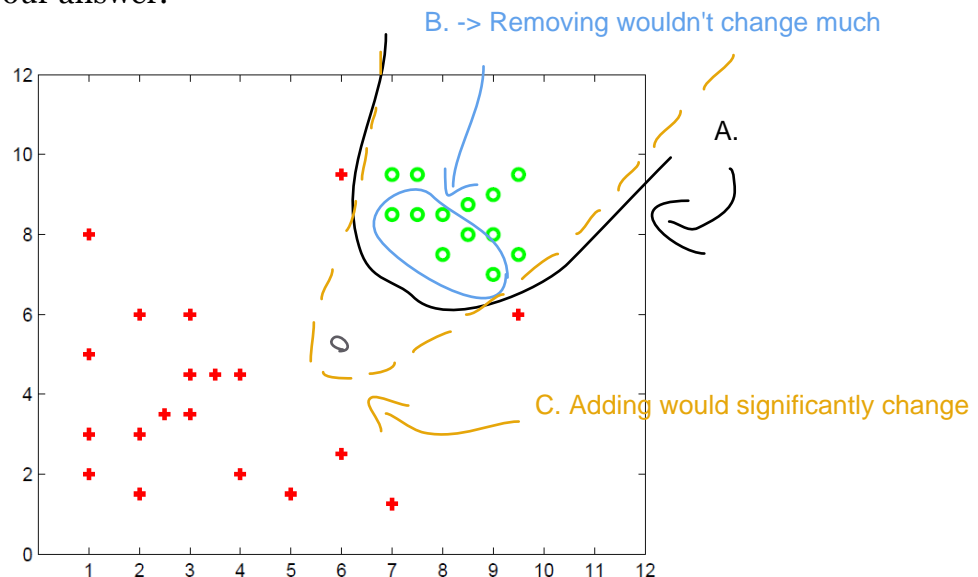


(b) (5 points) Mark a data point (among many) which will not change the decision boundary learned for very large values of $C$. Justify your answer.

(c) (5 points) Draw a data point which will significantly change the decision boundary learned for very large values of $C$. Justify your answer.

CSC 4850 /
6850

Practice
Midterm

2023/2
/23

(d) (7 points) For $C \to 0$, indicate in the figure below, where you would expect the decision boundary to be? Justify your answer.
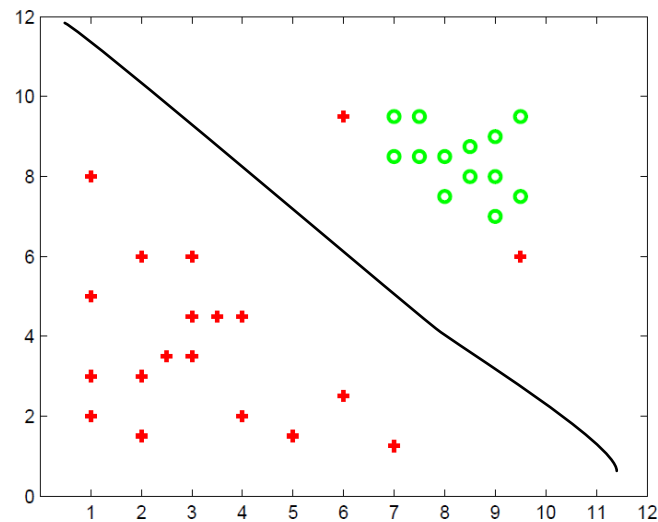


Didnt give penalty and just tried to maximize margin

(e) (6 points) Which of the two cases above (small $C$ or large $C$) would you expect to work better in the classification task? Why?
Small C because the two +'s by the o's are likely to be outliers