

CSE 404: Introduction to Machine Learning (Fall 2020)

Final Exam Take Home, Due: 11:59PM on Dec 16, 2020

- The exam should be completed independently and discussions of any type are NOT allowed.
- No late exam submission. D2L will be closed at 11:59pm and exams emailed after 11:59pm will NOT be accepted. Please plan early submissions to avoid Internet issues.

1. (10 points) Non-linear transformations

- (a) Please explain why non-linear transformations sometimes improve the classification performance of a linear model.
- (b) Below you are given two decision boundaries obtained for the same input. In Figure 1 (a), the decision boundary is obtained by logistic regression using the original data. In Figure 1 (b), the decision boundary is obtained by logistic regression using the data after applying a non-linear transformation. Which decision boundary would you prefer? Please explain the reason of your choice.

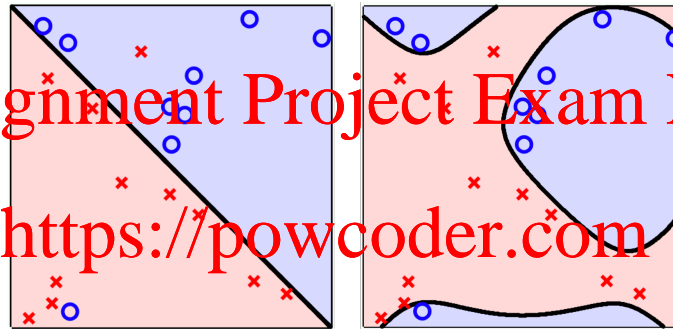


Figure 1: Decision boundaries learned by logistic regression.

2. (15 points) Support Vector Machines

Given two data points $x_1 = (1, 0)^T$, $y_1 = -1$, and $x_2 = (3, 0)^T$, $y_2 = 1$.

- (a) Compute the optimal w and b in support vector machine by solving the primal formulation given below:

$$\min_{w, b} \quad \frac{1}{2} w^T w, \quad \text{s.t.} \quad y_i(w^T x_i + b) \geq 1, \forall i.$$

- (b) Compute the optimal α in the dual formulation of support vector machine.
- (c) Compute the optimal w based on the optimal α obtained from the dual formulation of support vector machine. Compare with the results in (a). You should be able to obtain the same result.

3. (10 points) Support Vector Machines: Primal/Dual

- (a) Describe how the numbers of variables in the primal and dual formulations are related to the number of samples n and the data dimension d .

- (b) If you need to solve an SVM with Gaussian/RBF kernel on a data set with $n = 1,000,000$, $d = 1000$, which formulation (primal or dual) will you choose? Think about the number of parameters you will need to learn in primal and dual formulations and briefly explain your answer.
4. (15 points) **Support Vector Machines: Kernel Trick and Soft Margin**
- (a) Is $K(a, b) = (a^2 - b^4) - (a - b^2)^2$ a valid kernel function? Please explain.
 - (b) Explain how does the dual formulation of SVM allow using the kernel trick?
 - (c) Derive the dual soft-margin SVM formulation and clearly show your steps.
5. (10 points) **Loss Functions**
- (a) Write mathematical descriptions of hinge loss, logistic regression loss and square loss. Explain what each of the loss functions tries to achieve when we want to minimize them. *Hint: How each loss function behaves when there is a correct or wrong prediction.*
 - (b) Consider a correctly classified data point, which is very far away from the decision boundary. The decision boundaries obtained by SVM would not be affected by such a point, but the decision boundaries obtained by logistic regression would. Please explain why this is the case.
6. (10 points) **Unsupervised Learning**
- (a) What is the purpose of data standardization before computing the covariance matrix in PCA? Please explain.
 - (b) If we have supervision information about our data (binary or continuous labels), does PCA allow us to utilize this information?
7. (30 points) **Model Complexity**
- (a) What is model complexity? How is it related to sample size and regularization?
 - (b) Please discuss how to control model complexity in the context of algorithms we have learned (linear regression, logistic regression, decision tree, support vector machines).
 - (c) What is generalization performance and how does model complexity related to the generalization performance?