INF264 Introduction to machine learning
Spring 2020

Exam

10.2.2020

This exam has 5 tasks on 3 pages. You can get at most 50 points.
No aids permitted.

1 Basic concepts (10p)

Give short answers (about one paragraph) to the following questions:

1. What is overfitting and why is it a problem?
2. What is a kernel? Why is it useful?
3. Suppose that some values in our data matrix are missing. How can we handle this situation? Give at least two alternative ways.
4. Explain briefly how the k-nearest neighbor classifier works.
5. Consider k-means clustering. What is the goal? In other words, what is considered as a good cluster?

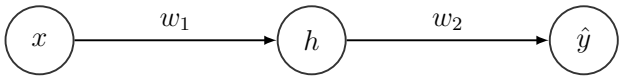
2 Boosting (10p)

1. What is boosting?
2. Write pseudocode for a basic boosting algorithm.
3. Why can boosting improve accuracy? Hint: use the bias-variance trade-off.

3 Neural networks (10p)

Consider the following simple neural network:

We have a one-dimensional input $x \in \mathbb{R}$ and a one-dimensional output $y \in \mathbb{R}$. Furthermore, we have one hidden layer consisting of one neuron and ReLU activation function. The output layer is linear.



That is, we have $z = w_1 x$, $h = f(z)$ where $f(z) = \max(0, z)$ and $\hat{y} = w_2 h$. We consider squared loss $L(y, \hat{y}) = \frac{1}{2}(y - \hat{y})^2$.

Suppose that initial weights have values $w_1 = 3$ and $w_2 = 2$. We have observed one data point with $x = 1$ and $y = 5$. Perform one update of parameters w_1 and w_2 using gradient descent with learning rate $\gamma = 0.1$. Show intermediate steps.

Hint: $f'(z) = 1$ when $z > 0$, $f'(z) = 0$ when $z < 0$ and undefined when $z = 0$.

4 PCA (10p)

Answer the following questions about principal component analysis (PCA).

1. Principal component analysis projects the data based on a criterion. What is this criterion?

2. How can we find the first principal component of the data?

3. Consider the 2-dimensional data shown in Figure 1. Draw the principal components.

4. Project the data to the first principal component. Draw the data points on the first principal component.

5. Project the points back to the original 2-dimensional space. Draw the points and the first principal component.

There is no need to get the drawings exactly right (and no need to draw all the points), it is enough to show that you have understood the idea.

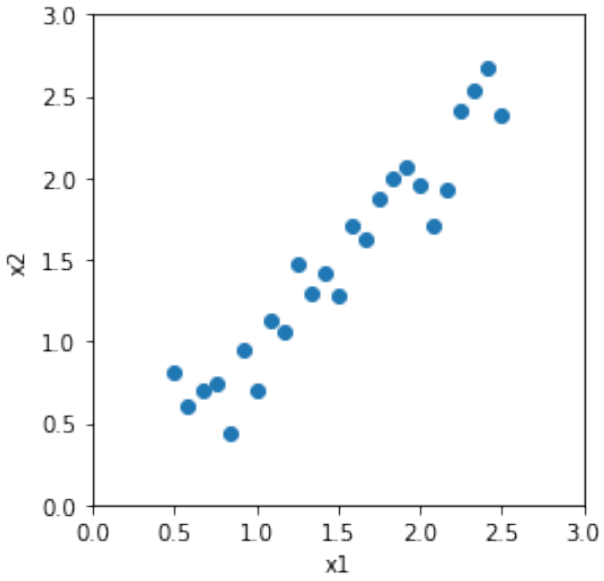


Figure 1: A data set for principal component analysis.

5 Model selection and evaluation (10p)

Suppose you are given 2000 datapoints $(\mathbf{x}_i, y_i)$ where $\mathbf{x}_i$ is a 1000-dimensional feature vector and $y_i \in \{0, 1\}$ is the class label. Consider k-nearest neighbor classifier with $k = 1, 3, 5$ and a neural network with $h = 1, 2, 3$ hidden layers (ReLU activation, 10 neurons per layer).

Explain a procedure that

1. selects the model with the smallest generalization error and
2. gives an unbiased estimate of the generalization error of the selected model.