

COMP9417 - Machine Learning

Tutorial: Nonparametric Modelling

Question 1. Expressiveness of Trees

Give decision trees to represent the following Boolean functions, where the variables A, B, C and D have values \mathbf{t} or \mathbf{f} , and the class value is either `True` or `False`. Can you observe any effect of the increasing complexity of the functions on the form of their expression as decision trees?

- (a) $A \wedge \neg B$
- (b) $A \vee [B \wedge C]$
- (c) $A \text{ XOR } B$
- (d) $[A \wedge B] \vee [C \wedge D]$

Question 2. Decision Tree Learning

- (a) Assume we learn a decision tree to predict class Y given attributes A, B and C from the following training set, with no pruning.

A	B	C	Y
0	0	0	0
0	0	1	0
0	0	1	0
0	1	0	0
0	1	1	0
0	1	1	1
1	0	0	0
1	0	1	1
1	1	0	1
1	1	0	1
1	1	1	0
1	1	1	1

What would be the training set error for this dataset? Express your answer as the number of examples out of twelve that would be misclassified.

- (b) One nice feature of decision tree learners is that they can learn trees to do *multi-class* classification, i.e., where the problem is to learn to classify each instance into exactly one of $k > 2$ classes.

Suppose a decision tree is to be learned on an arbitrary set of data where each instance has a discrete class value in one of $k > 2$ classes. What is the maximum training set error, expressed as a fraction, that any dataset could have?

Question 3. Linear Smoothing

In the lab this week we introduced linear smoothing, also known as kernel smoothing, and we implement it from scratch and apply it to a simulated dataset. The following figure is taken from [The Elements of Statistical Learning by Hastie, Tibshirani and Friedman](#) and is an excellent portrayal of the linear smoother at work. The data are simulated from the following data generating process:

$$y = \sin(4x) + \epsilon, \quad \epsilon \sim N(0, 1/3).$$

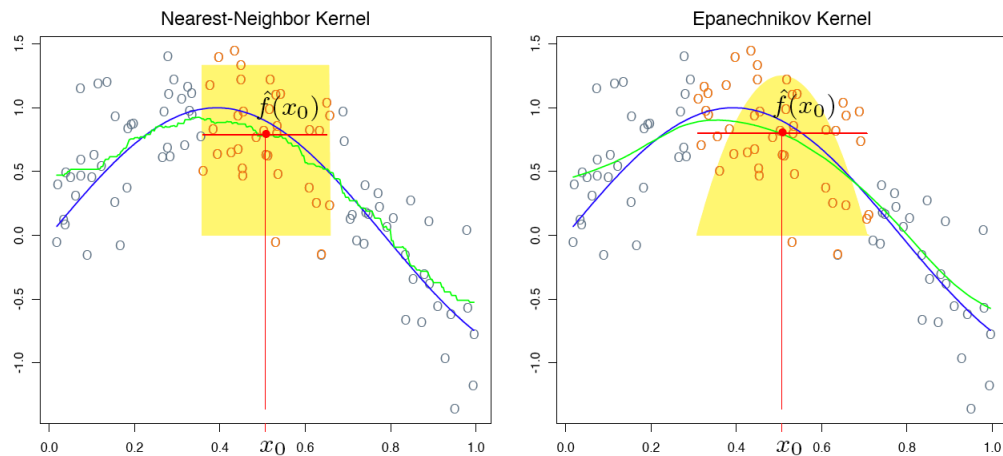
On the left, we see the ‘nearest-neighbour’ kernel at work, which we refer to as the boxcar kernel in the lab, and on the right we see the Epanechnikov kernel, also introduced in the lab. We include their definitions here:

$$K(u) = \mathbf{1}\{|u| \leq 1/2\} \quad \text{box-car kernel}$$

$$K(u) = \mathbf{1}\{|u| \leq 1\} \frac{3}{4}(1 - u^2) \quad \text{Epanechnikov kernel}$$

Recall also from the lab that a linear smoother prediction takes the form:

$$\hat{f}(x_0) = \sum_{i=1}^n \frac{K\left(\frac{\|X_i - x_0\|_2}{h}\right)}{\sum_{j=1}^n K\left(\frac{\|X_j - x_0\|_2}{h}\right)} Y_i.$$



Review the lab Linear smoothing section and then write down a few lines describing what is happening in the figure. Be sure to describe in detail what each of the following represents:

- (i) the blue curve
- (ii) the black scatter
- (iii) the red scatter

- (iv) the yellow region
- (v) the horizontal red line
- (vi) the red point on the horizontal red line
- (vii) the green curve