Classification and Regression

1. *Testing.* Consider a classification problem where we aim to classify new data points x into one of $c = 5$ classes denoted by $y = \{A, B, C, D, F\}$.

- *Training* in machine learning can be thought of as finding a function $f(x)$ that a good job of assigning each data x point to its correct class, $f(x)$. This is generally the challenging task in machine learning, and training is often computationally expensive.
- *Testing* simply refers to applying the function $f(\cdot)$ to a new data point x.

You have a function that predicts the final letter grade for this course based on the number of hours studied for the first exam, denoted x:

$$f(x) = \begin{cases} x \geq 10 & A \\ 10 > x \geq 8 & B \\ 8 > x \geq 6 & C \\ 6 > x \geq 4 & D \\ 4 > x \geq 0 & F. \end{cases}$$

a) John studies 5 hours for the first exam. What do you predict his final letter grade will be?

b) What is the set of possible outcomes $\mathcal{Y}$?

c) What is the set of input features $\mathcal{X}$?

d) You find that the classifier above does not work well. You decide to use additional information to build a better classifier. Specifically, you plan to make use of an additional binary indicator 0/1 which denotes if the student received a perfect score on homework 1. Specify the new set of possible inputs $\mathcal{X}$ (i.e, the feature set).

e) You learn (from looking at data) that a perfect score on homework 1 is approximately equivalent to 3.5 hours of studying. Specify a (reasonable) new classifier $f(\cdot)$.

2. Regression. Imagine that you want to build a function that predicts the high temperature (in degrees Celsius) in Madison based on the day of the year.

a) You start by considering a sinusoidal model of form:

$$f(x) = a + b\cos\left(\frac{2\pi x}{T} + \theta\right)$$

where $x \in \{1, 2, \ldots, 365\}$ indicates the day of the year. What is the set of possible outcomes $\mathcal{Y}$?

b) Imagine that you have some data from the last few years, of the form $\mathcal{D} = \{(\text{April 9 } 2001, 3.2), (\text{June 7 } 2014, 16), \ldots\}$. How can you estimate a, b, θ and T?

c) What is the fewest number of data points that you need to estimate a and b (assume that you know T)?

d) You next consider functions of the form:

$$f(x) = \begin{cases} x = 1 & t_1 \\ x = 2 & t_2 \\ & \vdots \\ x = 365 & t_{365} \end{cases}$$

where the values of t_i are learned from data. Without further assumptions, what is the fewest number of data points do you need to estimate the function?

e) What are some advantages or disadvantages of the two models?