Solutions.

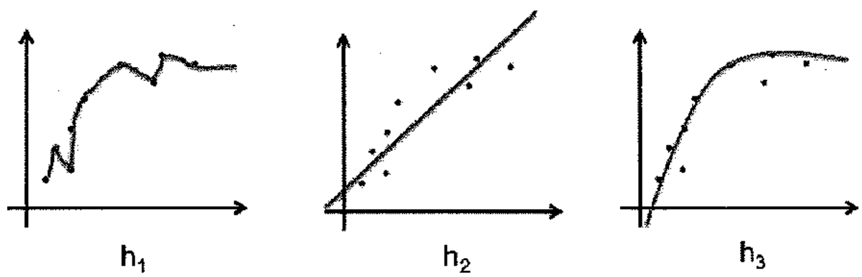
ECE 45234 - Spring 2018 Problem Set #4 Due: 4/25/18

The purpose of this problem set is to gain (admittedly limited) experience with learning in AI. In the engineering design section you will train and test a decision tree classifier..

You can complete the exercises by either directly marking up this pdf, or by printing, completing, and scanning as a pdf. You should complete the Engineering Design Problems by writing the python code as instructed. The resulting pdf and python files should be uploaded to Canvas via the assignment tab by the due date and time.

Exercises

1. Consider a regression problem. Given the following training data and three hypotheses (curves).



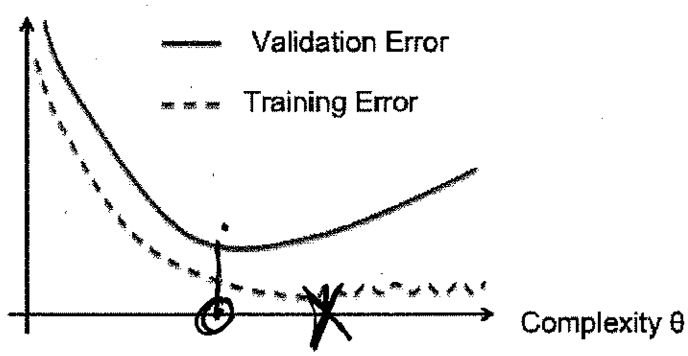
   (a) (2 points) Order the hypotheses according to their complexity. $h_2 < h_3 < h_1$

   (b) (2 points) Order the hypotheses according to their training error. $h_1 < h_3 < h_2$

   (c) (2 points) Which hypothesis do you think would generalize the best and why?

   h_1 clearly overfits, h_2 or h_3, likely h_3 would be chosen using cross-validation

2. Given the following training curves for an unspecified learning algorithm with complexity parameter θ,



   (a) (2 points) Mark with an X at what point cross-validation would stop training.

   (b) (2 points) Mark with an O what complexity cross-validation would select.

3. Suppose during training of a binary decision tree there are two attributes, A_1 with possible values $0, 1$ and A_2 with possible values $1, 2, 3$. There are 10 examples with the following class labels and attribute values

class	A_1	2
+1	0	2
+1	0	3
-1	1	2
+1	0	3
+1	1	2
-1	1	1
-1	1	1
-1	1	2
-1	0	1
-1	1	2

(a) (5 points) Compute the information gain for each attribute. Show your work.

Gain A_1

$v_0 = 0 \;\; p_0 = 3 \;\; n_0 = 1, \quad v_1 = 1 \;\; p_1 = 1, \; n_1 = 5, \quad p = 4, \; n = 6$

$Rem(A_1) = \frac{4}{10} B(\frac{3}{4}) + \frac{6}{10} B(\frac{1}{6}) = \frac{4}{10} 0.8 + \frac{6}{10} 0.65$

$Gain(A_1) = B(\frac{4}{10}) - Rem(A_1) = 0.256$

Gain A_2

$v_0 = 1 \;\; p_0 = 0 \;\; n_0 = 3 \;\; v_1 = 2 \;\; p_1 = 2, \; n_1 = 3 \;\; v_2 = 3 \;\; p_2 = 2 \;\; n_2 = 0$

$Rem(A_2) = \frac{3}{10} B(\frac{0}{3}) + \frac{5}{10} B(\frac{2}{5}) + \frac{2}{10} B(\frac{2}{2})$

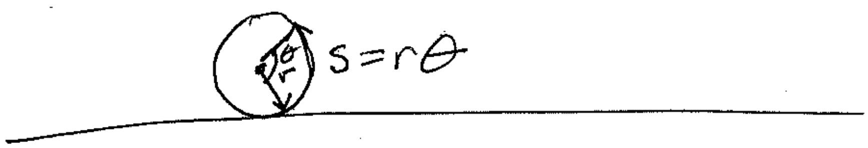
$Gain(A_2) = B(\frac{4}{10}) - Rem(A_2) = 0.485$

(b) (5 points) Which attribute would be selected first and why?

Choose A_2 because $Gain(A_2) > Gain(A_1)$

4. Suppose a rover has an electric motor driving a wheel and a sensor that measures the rotation angle of the motor. The robot needs to learn a function that maps the measured rotation angle to a distance translated by the rover. Because of variations in manufacturing the radius of the wheel is unknown perfectly, but the specifications say it's mean is 3cm and has a standard deviation of 3mm.

(a) (5 points) How could one cast this calibration problem as a learning problem?



Let y be the distance traveled when X = anghe.

Because of slippage. $y = rX + \eta$ where

η is a R.V. with known statistics.

We assume X is known, thus we want to find given measured Data $D = \{y_i\}$

$$p(r/D) \propto P(D/r)\,p(r) \quad p(r) \sim N(3,3)$$

(b) (5 points) How could one design the robot to learn the mapping from rotation to translation?

You need to perform a calibration, where preprogramed angles are produced and distances measured. to Get a trainy set D.

Then use MAP learning to combine the Data and prior to estimate r, the wheel radius.