Comp 6321 - Machine Learning - Assignment 3

Federico O'Reilly Regueiro

November 10^{th}, 2016

Question 1: Midterm preparation question

Propose an adequate learning algorithm for each instance.

1.a 1000 samples, 6-dimensional continuous space, classify ~100 examples.

1.b Clasifier for children in special-ed, justified to the board before it's implemented.

1.c Binary classification of 1 million bits (empirical preference rate for others), very large data-set. Frequent updates.

1.d 40 attributes, discrete and continuous, some have noise; only about 50 labeled observations.

Question 2: Properties of entropy

2.a Compute the following for (X, Y): $p(0,0) = 1/3, p(0,1) = 1/3, p(1,0) = 0, p(1,1) = 1/3$.

i $H[x] = -\frac{1}{3}log_2\left(\frac{1}{3}\right) - \frac{2}{3}log_2\left(\frac{2}{3}\right) = .9182$

ii $H[y] = -\frac{1}{3}log_2\left(\frac{1}{3}\right) - \frac{2}{3}log_2\left(\frac{2}{3}\right) = .9182$

iii $H[y|x] = \sum_x p(x)H[Y|X=x] = \frac{2}{3}\left(-\frac{1}{2}log_2\left(\frac{1}{2}\right) - \frac{1}{2}log_2\left(\frac{1}{2}\right)\right) = \frac{2}{3}$

iv $H[x|y] = \sum_y p(x)H[X|Y=y] = \frac{2}{3}\left(-\frac{1}{2}log_2\left(\frac{1}{2}\right) - \frac{1}{2}log_2\left(\frac{1}{2}\right)\right) = \frac{2}{3}$

v $H[x,y] = 3\left(-\frac{1}{3}log_2\left(\frac{1}{3}\right)\right) = -log_2\left(\frac{1}{3}\right) = 1.5849$

vi $I[x,y] = \sum_x \sum_y p(x,y)log_2\left(\frac{p(x,y)}{p(x)p(y)}\right) = H[x] - H[x|y] = 0.2516$

2.b Prove maximum entropy in a discrete distribution happens in U

We wish to find:

$$\arg\max_{p_n} \sum_{n=1}^{N} p_n log(p_n)$$

With constraints:

$$1 - \sum_{n=1}^{N} p_n = 0$$

We use a Lagrangian multiplier such that:

$$\nabla_{p_1,p_2,\dots p_N} \sum_{n=1}^{N} p_n log(p_n) = \nabla_{p_1,p_2\dots p_N} \lambda(1 - \sum_{n=1}^{N} p_n)$$

We are thus left with a system:

$$\begin{aligned}
\frac{\partial}{\partial_{p_1}} \sum_{n=1}^{N} p_n log(p_n) =& \frac{\partial}{\partial_{p_1}} \lambda(1 - \sum_{n=1}^{N} p_n) \\
\frac{\partial}{\partial_{p_2}} \sum_{n=1}^{N} p_n log(p_n) =& \frac{\partial}{\partial_{p_2}} \lambda(1 - \sum_{n=1}^{N} p_n) \\
&\vdots \\
\frac{\partial}{\partial_{p_N}} \sum_{n=1}^{N} p_n log(p_n) =& \frac{\partial}{\partial_{p_N}} \lambda(1 - \sum_{n=1}^{N} p_n) \\
1 - \sum_{n=1}^{N} p_n =& 0
\end{aligned}$$

Which in turn yields:

$$\begin{aligned}
log(p_1) + 1 =& \lambda p_1 \\
log(p_2) + 1 =& \lambda p_2 \\
&\vdots \\
log(p_N) + 1 =& \lambda p_N \\
1 - \sum_{n=1}^{N} p_n =& 0
\end{aligned}$$

From which it is clear that $p_1 = p_2 = \dots p_N = \frac{1}{N}$, which is precisely a discrete uniform distribution.

2.c Show that T_1 wins

The notes show two possible tests for a decision tree. T1, where the left child has $[20+, 10-]$ posible outcomes in its sub-trees and the right node has $[10+, 0-]$. T2, on the other hand, yields: $left = [15+, 7-]; right = [15+, 3-]$.

The best choice should yield the maximum information gain $I[p, T_n]$,$n \in \{1, 2\}$. So for T_1:

$$
\begin{aligned}
H[p] =& -\frac{1}{4}log_2\left(\frac{1}{4}\right) - \frac{3}{4}log_2\left(\frac{3}{4}\right) = 0.8112 \\
H[p|T_1 = t] =& -\frac{2}{3}log_2\left(\frac{2}{3}\right) - \frac{1}{3}log_2\left(\frac{1}{3}\right) = 0.9182 \\
H[p|T_1 = f] =& 0 \\
H[p|T_1] =& p(T_1 = t)H[p|T_1 = t] + p(T_1 = f)H[p|T_1 = f] \\
=& 0.6887 \\
I[p, T_1] =& H[p] - H[p|T_1] = 0.1225
\end{aligned}
$$

Whereas for T_2 we have:

$$
\begin{aligned}
H[p|T_2 = t] =& -\frac{15}{22}log_2\left(\frac{15}{22}\right) - \frac{7}{22}log_2\left(\frac{7}{22}\right) = 0.9024 \\
H[p|T_2 = f] =& -\frac{15}{18}log_2\left(\frac{15}{18}\right) - \frac{3}{18}log_2\left(\frac{3}{18}\right) = 0.65002 \\
H[p|T_2] =& p(T_2 = t)H[p|T_2 = t] + p(T_2 = f)H[p|T_2 = f] \\
=& \frac{22}{40}0.9024 + \frac{18}{40}0.65002 = 0.7888 \\
I[p, T_2] =& H[p] - H[p|T_2] = 0.02245
\end{aligned}
$$

From which we can see that we gain much more information from knowing the result of T_1 than by knowing the result of T_2.

Question 3: Kernels

3.a

3.b

3.c

3.d

3.e

Question 4: Nearest neighbour vs decision trees

Question 5: Bayes rate

5.a

5.b

5.c

Question 6: Implementation