Ryan Karl
Data Science FA19-CSE-40647-CX-01
October 7, 2019
Dr. Meng Jiang

Homework 2 Solutions

Solution 1:

Part A:

Prior Probability:

$P(Y = 0) = \frac{3}{6} = 0.5$ and $P(Y = 1) = \frac{3}{6} = 0.5$.

Likelihood Probability:

$P(A = 1 \mid Y = 1) = \frac{2}{3}$ and $P(A = 1 \mid Y = 0) = \frac{1}{3}$

$P(B = 0 \mid Y = 1) = \frac{1}{3}$ and $P(B = 0 \mid Y = 0) = \frac{1}{3}$

$P(C = 0 \mid Y = 1) = \frac{2}{3}$ and $P(C = 0 \mid Y = 0) = \frac{1}{3}$

$P(X|Y = 1) = (\frac{2}{3}) * (\frac{1}{3}) * (\frac{2}{3}) = \frac{4}{27}$ and $P(X|Y = 0) = (\frac{1}{3}) * (\frac{1}{3}) * (\frac{1}{3}) = \frac{1}{27}$

Posteriori Probability:

$P(Y = 1|X) = (\frac{4}{27}) * (\frac{1}{2}) = (\frac{2}{27})$ and $P(Y = 0|X) = (\frac{1}{27}) * (\frac{1}{2}) = (\frac{1}{54})$

We predict that Y=1 with Naive Bayes.

Part B:

Step 1:

$$Y = \{1 * 3, 0 * 3\}$$

$$H(Y) = -(\frac{1}{2})log(\frac{1}{2}) - (\frac{1}{2})log(\frac{1}{2}) = 0.301$$

$$X_A = \{1 * 3, 0 * 3\}$$

$$H(Y \mid X_A) = H(Y \mid A_1) + H(Y \mid A_0) = (\frac{1}{2})* (-(\frac{2}{3})\log(\frac{2}{3}) - (\frac{1}{3})\log(\frac{1}{3})) + (\frac{1}{2})*(-(\frac{1}{3})\log(\frac{1}{3})-(\frac{2}{3})\log(\frac{2}{3})) = 0.276$$

$$IG(Y|X_A) = 0.301 - 0.276 = 0.025$$

$$X_B = \{1 * 4, 0 * 2\}$$

$$H(Y \mid X_B) = H(Y \mid B_1) + H(Y \mid B_0) = (\frac{2}{3})* (-(\frac{1}{2})\log(\frac{1}{2}) - (\frac{1}{2})\log(\frac{1}{2})) + (\frac{1}{3})*(-(\frac{1}{2})\log(\frac{1}{2})-(\frac{1}{2})\log(\frac{1}{2})) = 0.301$$

$$IG(Y|X_C) = 0.301 - 0.301 = 0$$

$$X_C = \{1 * 3, 0 * 3\}$$

$$H(Y \mid X_C) = H(Y \mid C_1) + H(Y \mid C_0) = (\frac{1}{2})* (-(\frac{1}{3})\log(\frac{1}{3}) - (\frac{2}{3})\log(\frac{2}{3})) + (\frac{1}{2})*(-(\frac{2}{3})\log(\frac{2}{3})-(\frac{1}{3})\log(\frac{1}{3})) = 0.276$$

$$IG(Y|X_C) = 0.301 - 0.276 = 0.025$$

Note that we could have chosen C instead of A since the information gain for both is the same.

Step 2a:

$$Y = \{1 * 2, 0 * 1\}$$

$$H(Y) = -(\frac{1}{3})log(\frac{1}{3}) - (\frac{2}{3})log(\frac{2}{3}) = 0.276$$

$$X_B = \{1 * 2, 0 * 1\}$$

$$\text{H}(\text{Y} \mid X_B) = \text{H}(\text{Y} \mid B_1) + \text{H}(\text{Y} \mid B_0) = (\frac{2}{3})^* (-(\frac{1}{2})\log(\frac{1}{2}) - (\frac{1}{2})\log(\frac{1}{2})) + (\frac{1}{3})^*(-(\frac{1}{1})\log(\frac{1}{1})) = 0.200$$

$$IG(Y|X_B) = 0.276 - 0.200 = 0.076$$

$$X_C = \{1 * 2, 0 * 1\}$$

$$\text{H}(\text{Y} \mid X_C) = \text{H}(\text{Y} \mid C_1) + \text{H}(\text{Y} \mid C_0) = (\frac{2}{3})^* (-(\frac{1}{2})\log(\frac{1}{2}) - (\frac{1}{2})\log(\frac{1}{2})) + (\frac{1}{3})^*(-(\frac{1}{1})\log(\frac{1}{1}))= 0.200$$

$$IG(Y|X_C) = 0.276 - 0.200 = 0.076$$

Note that we could have chosen C instead of B since the information gain for both is the same.

Step 2b:

$$Y = \{1 * 1, 0 * 2\}$$

$$H(Y) = -(\frac{1}{3})log(\frac{1}{3}) - (\frac{2}{3})log(\frac{2}{3}) = 0.276$$

$$X_B = \{1 * 2, 0 * 1\}$$

$$\text{H}(\text{Y} \mid X_B) = \text{H}(\text{Y} \mid B_1) + \text{H}(\text{Y} \mid B_0) = (\frac{2}{3})^* (-(\frac{1}{2})\log(\frac{1}{2}) - (\frac{1}{2})\log(\frac{1}{2})) + (\frac{1}{3})^*(-(\frac{1}{1})\log(\frac{1}{1})) = 0.200$$

$$IG(Y|X_B) = 0.276 - 0.200 = 0.076$$

$$X_C = \{1 * 1, 0 * 2\}$$

$$\text{H}(\text{Y} \mid X_C) = \text{H}(\text{Y} \mid C_1) + \text{H}(\text{Y} \mid C_0) = (\frac{1}{3})^* (-(\frac{1}{1})\log(\frac{1}{1}) + (\frac{2}{3})^*(-(\frac{1}{2})\log(\frac{1}{2}) \;-\; (\frac{1}{2})\log(\frac{1}{2}))= 0.200$$

$$IG(Y|X_C) = 0.276 - 0.200 = 0.076$$

Note that we could have chosen C instead of B since the information gain for both is the same.

Step 3a:

$$Y = \{1 * 1, 0 * 1\}$$

$$H(Y) = -(\frac{1}{2})log(\frac{1}{2}) - (\frac{1}{2})log(\frac{1}{2}) = 0.301$$

$$X_C = \{1 * 1, 0 * 1\}$$

$$\text{H}(\text{Y} \mid X_B) = \text{H}(\text{Y} \mid B_1) + \text{H}(\text{Y} \mid B_0) = (\frac{1}{2})^* (-(\frac{1}{1})\log(\frac{1}{1}) - (\frac{1}{1})\log(\frac{1}{1})) + (\frac{1}{2})^*(-(\frac{1}{1})\log(\frac{1}{1})) = 0.00$$

$$Pure\ branches\ (\ nothing\ learned\ via\ split)\ so\ if\ C = 0\ then\ Y = 1\ and\ if\ C = 1\ then\ Y = 0$$

Step 3b:

$$Y = \{1 * 1, 0 * 1\}$$

$$H(Y) = -\left(\frac{1}{1}\right)\log\left(\frac{1}{1}\right) - \left(\frac{0}{1}\right)\log\left(\frac{0}{1}\right) = 0$$

$$Pure\ branch\ (\ nothing\ learned\ via\ split)\ so\ if\ C = 0\ or\ C = 1\ then\ Y = 1$$

Step 3c:

$$Y = \{1 * 1, 0 * 1\}$$

$$H(Y) = -(\frac{1}{2})log(\frac{1}{2}) - (\frac{1}{2})log(\frac{1}{2}) = 0.301$$

$$X_C = \{1 * 1, 0 * 1\}$$

$$\mathrm{H}(\mathrm{Y} \mid X_B) = \mathrm{H}(\mathrm{Y} \mid B_1) + \mathrm{H}(\mathrm{Y} \mid B_0) = (\tfrac{1}{2})* \left(-(\tfrac{1}{1})\log(\tfrac{1}{1}) - (\tfrac{1}{1})\log(\tfrac{1}{1})\right) + (\tfrac{1}{2})*\left(-(\tfrac{1}{1})\log(\tfrac{1}{1})\right) = 0.00$$

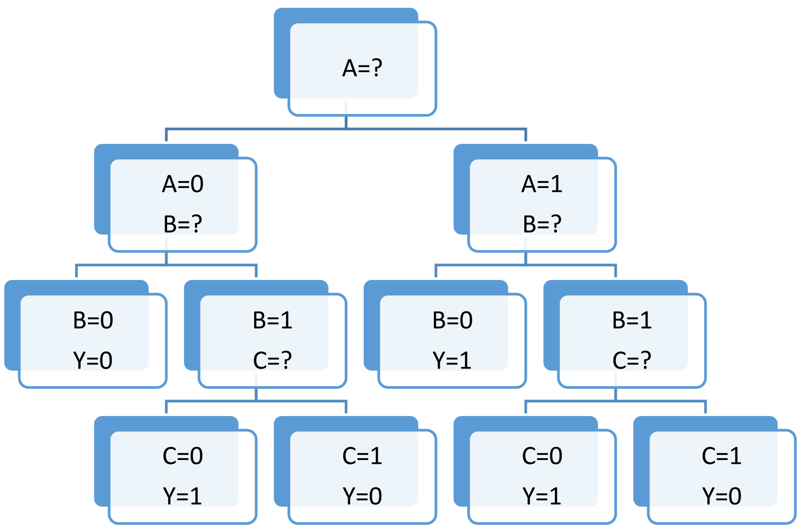
$$\textit{Pure branches (nothing learned via split) so if } C = 0 \textit{ then } Y = 1 \textit{ and if } C = 1 \textit{ then } Y = 0$$

Step 3d:

$$Y = \{1 * 1, 0 * 1\}$$

$$H(Y) = -\left(\frac{1}{1}\right)\log\left(\frac{1}{1}\right) - \left(\frac{0}{1}\right)\log\left(\frac{0}{1}\right) = 0$$

$$\textit{Pure branch (nothing learned via split) so if } C = 0 \textit{ or } C = 1 \textit{ then } Y = 0$$



__**Solution 2:**__

We have $K(x_i, x_j) = (x_i x_j + 1)^3$. Note that for $x = (x_1, x_2)$ this is $(x_{i1}x_{j1} + x_{i2}x_{j2} + 1)^3$, so:

$$K(x_i, x_j) = (x_i x_j + 1)^3$$

$$= (x_{i1}x_{j1}x_{i1}x_{j1} + x_{i1}x_{j1}x_{i2}x_{j2} + x_{i1}x_{j1} + x_{i2}x_{j2}x_{i1}x_{j1} + x_{i2}x_{j2}x_{i2}x_{j2} + x_{i2}x_{j2} + x_{i1}x_{j1} + x_{i2}x_{j2} + 1)(x_{j1}x_{i1} + x_{i2}x_{j2} + 1)$$

$$=({x_{j1}}^2{x_{i1}}^2 + 2x_{j1}x_{i1} + 2x_{j1}x_{j2}x_{i1}x_{i2} + {x_{j2}}^2{x_{i2}}^2 + 2x_{j2}x_{i2} + 1)(x_{j1}x_{i1} + x_{j2}x_{i2} + 1)$$

$$= {x_{j1}}^2{x_{i1}}^2x_{j1}x_{i1} + {x_{j1}}^2{x_{i1}}^2x_{i2}x_{j2} + {x_{j1}}^2{x_{i1}}^2 + 2x_{j1}x_{i1}x_{j1}x_{i1} + 2x_{j1}x_{i1}x_{i2}x_{j2} + 2x_{j1}x_{i1} + 2x_{j1}x_{j2}x_{i2}x_{i1}x_{i1}x_{j1} + 2x_{j1}x_{j2}x_{i2}x_{i1}x_{i2}x_{j2} + 2x_{j1}x_{j2}x_{i1}x_{i2} + {x_{j2}}^2{x_{i2}}^2x_{i1}x_{j1} + {x_{j2}}^2{x_{i2}}^2x_{j2}x_{i2} + {x_{j2}}^2{x_{i2}}^2 + 2x_{j2}x_{i2}x_{i1}x_{j1} + 2x_{j2}x_{i2}x_{i2}x_{j2} + 2x_{j2}x_{i2} + x_{i1}x_{j1} + x_{j2}x_{i2} + 1$$

$$= {x_{j1}}^3{x_{i1}}^3 + 3{x_{j1}}^2{x_{i1}}^2 + 3{x_{j1}}^2x_{j2}{x_{i1}}^2x_{i2} + 3x_{j1}x_{i1} + 3x_{j1}{x_{j2}}^2x_{i1}{x_{i2}}^2 + 6x_{j1}x_{j2}x_{i1}x_{i2} + {x_{j2}}^3{x_{i2}}^3 + 3{x_{j2}}^2{x_{j2}}^2 + 3x_{j2}x_{i2} + 1$$

So then:

$$\Phi(x) = x_1^3 + \sqrt{3}x_1^2 + \sqrt{3}x_1^2x_2 + \sqrt{3}x_1 + \sqrt{3}x_1x_2^2 + \sqrt{6}x_1x_2 + x_2^3 + \sqrt{3}x_2^2 + \sqrt{3}x_2 + 1$$

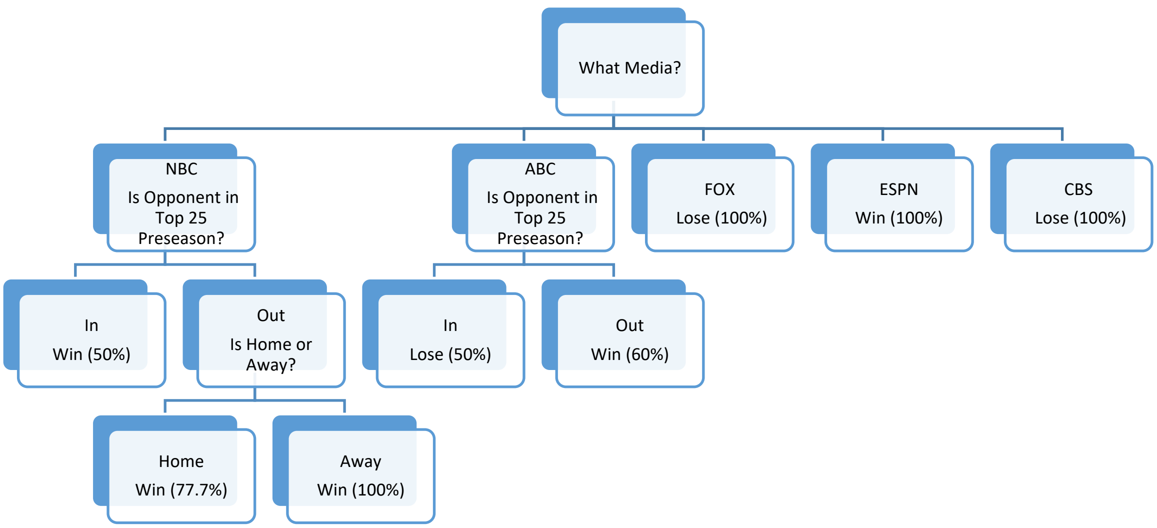
Solution 3:

Note that if $TP = TN$ we have that Accuracy $= \frac{TP+TN}{Total\ Population} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{2TP}{2TP+FP+FN}$. Now $Precision = \frac{TP}{TP+FP}$ and $Recall = \frac{TP}{TP+FN}$, so if $TP = TN$ we have that $F_1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} =$

$$2 \times \frac{(\frac{TP}{TP+FP}) \times (\frac{TP}{TP+FN})}{(\frac{TP}{TP+FP}) + (\frac{TP}{TP+FN})} = 2 \times \frac{(\frac{(TP)^2}{(TP+FP)\times(TP+FN)})}{(\frac{(2(TP)^2)+(TP\times FP)+(TP\times FN)}{(TP+FP)\times(TP+FN)})} = 2 \times (\frac{(TP)^2}{(TP+FP)\times(TP+FN)}) \times (\frac{(TP+FP)\times(TP+FN)}{(2(TP)^2)+(TP\times FP)+(TP\times FN)}) = 2 \times \left(\frac{(TP)^2}{(2(TP)^2)+(TP\times FP)+(TP\times FN)}\right) = 2 \times \left(\frac{(TP)^2}{(TP)\times((2TP)+(FP)+(FN))}\right) = 2 \times \left(\frac{TP}{((2TP)+(FP)+(FN))}\right) = \frac{2TP}{2TP+FP+FN}$$

Since both Accuracy and F_1 can be derived to equivalent expressions we know they are equivalent when $TP = TN$.

Solution 1:

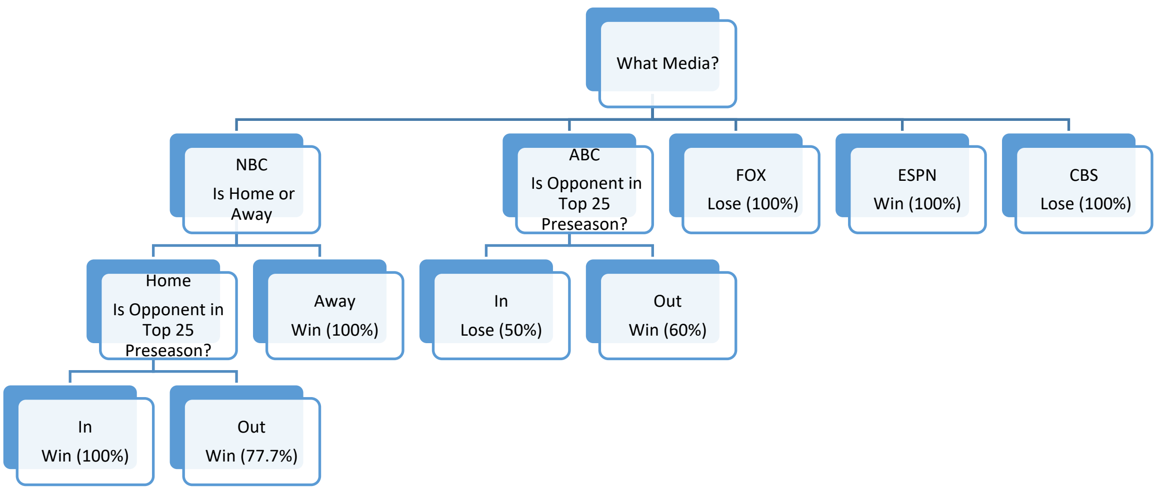


Game #	Predicted Label
1	Win
2	Win
3	Win
4	Lose
5	Win
6	Win
7	Win
8	Win
9	Win
10	Lose
11	Win

12	Lose

Metric	Value
Accuracy	0.833333333333
Precision	0.888888888889
Recall	0.888888888889
F1 Score	0.444444444444

Solution 2:



Game #	Predicted Label
1	Win
2	Win
3	Win
4	Lose
5	Win
6	Win
7	Win
8	Win
9	Win
10	Lose
11	Win
12	Lose

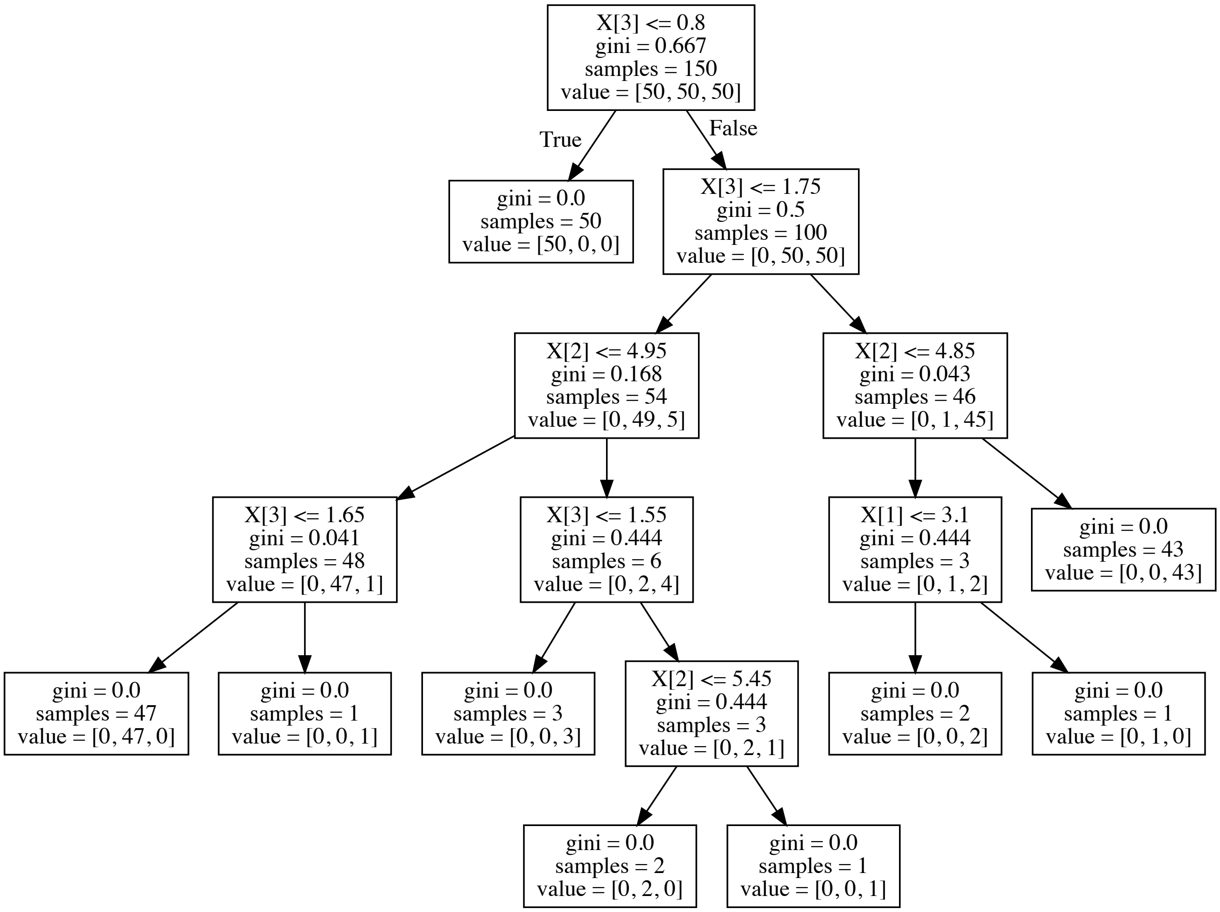
Metric	Value
Accuracy	0.833333333333
Precision	0.888888888889
Recall	0.888888888889
F1 Score	0.444444444444

Solution 3:

Game #	Predicted Label
1	Win
2	Win
3	Win
4	Lose
5	Win
6	Lose
7	Win
8	Win
9	Win
10	Lose
11	Win
12	Lose

Metric	Value
Accuracy	0.75
Precision	0.875
Recall	0. 777777777778
F1 Score	0. 411764705882

Solution 4:


X[3] <= 0.8
gini = 0.667
samples = 150
value = [50, 50, 50]
True
False
gini = 0.0
samples = 50
value = [50, 0, 0]
X[3] <= 1.75
gini = 0.5
samples = 100
value = [0, 50, 50]
X[2] <= 4.95
gini = 0.168
samples = 54
value = [0, 49, 5]
X[2] <= 4.85
gini = 0.043
samples = 46
value = [0, 1, 45]
X[3] <= 1.65
gini = 0.041
samples = 48
value = [0, 47, 1]
X[3] <= 1.55
gini = 0.444
samples = 6
value = [0, 2, 4]
X[1] <= 3.1
gini = 0.444
samples = 3
value = [0, 1, 2]
gini = 0.0
samples = 43
value = [0, 0, 43]
gini = 0.0
samples = 47
value = [0, 47, 0]
gini = 0.0
samples = 1
value = [0, 0, 1]
gini = 0.0
samples = 3
value = [0, 0, 3]
X[2] <= 5.45
gini = 0.444
samples = 3
value = [0, 2, 1]
gini = 0.0
samples = 2
value = [0, 0, 2]
gini = 0.0
samples = 1
value = [0, 1, 0]
gini = 0.0
samples = 2
value = [0, 2, 0]
gini = 0.0
samples = 1
value = [0, 0, 1]