Sta 671 Hw 1.

1. 1.

Regression

2. Regression

3. Classification

4. Regression

5. Classification.

6. pattern mining.

7. clustering.

8. Pattern mining.

9. density estimation

10. Conditional probability estimation

11. Regression / classification.

12. ranking.

13. Pattern mining.

3. Entropy: $H(Y|X) = \sum_x H(Y|X=x)\, P(X=x)$

Gain: $G(X,Y) := I(X,Y) := H(Y) - H(Y|X)$

3.1

$$G(X,X) = I(X,X) = H(X) - H(X|X)$$

$$H(X|X) = -\sum_{i,j} p(x_i, x_j) \log \frac{p(x_i, x_j)}{p(x_i)}$$

$$= -\sum_{i,j} p(x_i, x_j) \log \frac{p(x_i)}{p(x_i)}$$

$$= \sum_{i,j} p(x_i, x_j) \cdot \log(1)$$

$$= 0$$

Then, $I(X,X) := H(X) - H(X|X)$

$$= H(X) - 0$$

$$= H(X).$$

3.2. $I(x) = -\log(x)$

$$I(\Pi E_i) = -\log(\Pi E_i)$$

$$= -\log(E_1) - \log(E_2) - \cdots - \log(E_n)$$

$$= -\sum \log(E_i)$$

$$= \sum I(E_i).$$

3.3. w.t.s. $I(X|Y) \geq 0$.

$\Leftrightarrow H(Y) - H(Y|X) \geq 0$

$$H(Y) - H(Y|X)$$

$$= \sum_{y} P(Y=y) \log\left(\frac{1}{P(Y=y)}\right) - \sum_{x,y} P(X=x) P(Y=y|X=x) \cdot \log \frac{1}{P(Y=y|X=x)}$$

$$= \sum_{y} P(Y=y) \log\left(\frac{1}{P(Y=y)}\right) \cdot \sum_{x} P(X=x|Y=y) - \sum_{x,y} P(X=x) P(Y=y|X=x) \cdot \log \frac{1}{P(Y=y|X=x)}$$

Since $\sum P(Y=y|X=x) = 1$

$$= \sum_{x,y} P(X=x, Y=y) \cdot \log\left(\frac{P(Y=y|X=x)}{P(Y=y)}\right)$$

$$= \sum_{x,y} P(X=x, Y=y) \log\left(\frac{P(X=x, Y=y)}{P(Y=y)\,P(X=x)}\right)$$

$$= -\sum_{x,y} P(X=x, Y=y) \log\left(\frac{P(Y=y)\,P(X=x)}{P(X=x, Y=y)}\right)$$

By Jensen's Inequality of convex function.

$$\geq -\log\left(\sum_{x,y} P(X=x, Y=y) \cdot \frac{P(Y=y) \cdot P(X=x)}{P(X=x, Y=y)}\right)$$

$$\geq -\log\left(\sum_{x,y} P(X=x) P(Y=y)\right)$$

$$\geq -\log(1)$$

$$\geq 0.$$

Thus, $H(Y) - H(Y|X) \geq 0$, and $I(X;Y) \geq 0$.

3.4. $Z = X + Y$ $\quad Y = Z - X$, $\quad \frac{dy}{dz} = 1$.

a/. $H(Z|X) = \sum p(x) H(Z|X=x)$

$= -\sum_x p(x) \sum_z p(Z=z|X=x) \cdot \log p(Z=z|X=x)$

by $Z = X+Y$.

$= -\sum_x p(x) \sum_y p(Y = Z-x \,|\, X=x) \log p(\underbrace{Y = Z-x}_{y} \,|\, X=x)$

$= \sum_x p(x) \; H(Y|X=x)$

$= H(Y|X)$

b/. Given $X \perp Y$.

By 3.3, $I(X,Z) \geq 0$. $\Rightarrow H(Z) \geq H(Z|X) = H(Y|X)$

$H(Y|X) = \sum_x H(Y|X=x) P(X=x)$

$= -\sum_x P(Y|X) \log(P(Y|X=x))$

$= -\sum_x \frac{P(X,Y)}{P(X)} \log\left(\frac{P(X,Y)}{P(X)}\right)$

Since $X \perp Y$,

$= -\sum_x P(Y) \log(P(Y))$

$= H(Y)$.

Then, $H(Z) \geq H(Y)$.

Similarly, $H(Z|Y) = H(X|Y)$. by a similar proof to prove a.

Then, $H(Z) \geq H(Z|Y)$ since $I(Y, Z) \geq 0$.

$\Rightarrow H(Z) \geq H(X|Y)$

By a similar proof, we have $H(X|Y) = H(X)$ as $X \perp Y$.

$\Rightarrow H(Z) \geq H(X)$.

Thus, if $X \perp Y$, $H(Z) \geq H(Y)$ and $H(Z) \geq H(X)$.

4.

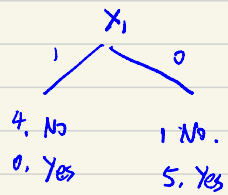
4.1

a).

Gini Index $(P, 1-P) = 2 \cdot 0.5 \cdot (1 - 0.5)$

$= 0.5.$

If split on X_1



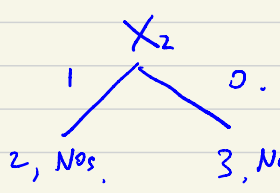
Gini-reduction $= 0.5 - 0 - 2 \cdot \frac{1}{6} \cdot \frac{5}{6} \cdot \frac{6}{10}$.

$= \frac{1}{2} - \frac{5}{18} \cdot \frac{6}{10}$

$= \frac{1}{2} - \frac{1}{6}$

$= \frac{1}{3}$.

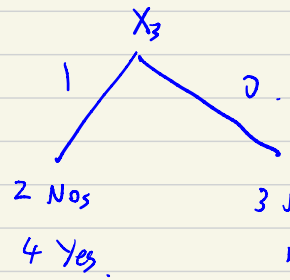
If we split on X_2.



Gini Reduction on $X_2 = \frac{1}{2} - 0.5 \cdot \frac{2}{5} \cdot \frac{3}{5} \cdot 2 - 0.5 \cdot \frac{3}{5} \cdot \frac{2}{5} \cdot 2.$

$= \frac{1}{2} - 2 \cdot \frac{2}{5} \cdot \frac{3}{5} = 0.02.$

If we split on X_3.



Gini Reduction on $X_2 = \frac{1}{2} - \frac{6}{10} \cdot 2 \cdot \frac{2}{3} \cdot \frac{1}{3} - \frac{4}{10} \cdot 2 \cdot \frac{3}{4} \cdot \frac{1}{4}$

$= \frac{1}{2} - \frac{4}{15} - \frac{3}{20} = 0.083.$

Since split on X_1 first has the best Gini-Reduction, we would thus pick X_1 as the first feature to split, based on Gini-Index criteria.

b/ $H(Y) = H([0.5, 0.5])$

$= -2 \cdot \frac{1}{2} \log_2 \frac{1}{2} = 1.$

If split on X_1

4. No
0. Yes

1 No.
5. Yes

Information Gain on X_1 =

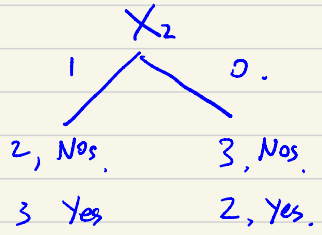
$1 - \left(\frac{4}{10} \cdot H([1, 0]) + \frac{6}{10} \cdot H\left([\frac{1}{6}, \frac{5}{6}]\right) \right)$

$= 1 - \frac{4}{10} \cdot 0 + \frac{6}{10} \cdot \left(\frac{1}{6} \log_2 \left(\frac{1}{6}\right) + \frac{5}{6} \log_2 \left(\frac{5}{6}\right) \right)$

$= 1 - 0.65$

$= 0.35.$

If we split on X_2.



Info. Gain on X_2:

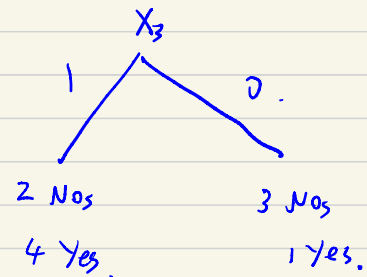
$$1 - \left(0.5\, H\left(\left[\tfrac{2}{5}, \tfrac{3}{0}\right]\right) + 0.5\, H\left(\left[\tfrac{2}{5}, \tfrac{3}{5}\right]\right)\right)$$

$$= 1 - H\left(\left[\tfrac{2}{5}, \tfrac{3}{5}\right]\right)$$

$$= 1 - 0.97$$

$$= 0.03$$

If we split on X_3.



Information Gain on splitting X_3:

$$1 - \frac{6}{10} \cdot H\left(\left[\tfrac{1}{3}, \tfrac{2}{3}\right]\right) - \frac{4}{10} H\left(\left[\tfrac{3}{4}, \tfrac{1}{4}\right]\right)$$

$$= 0.124.$$

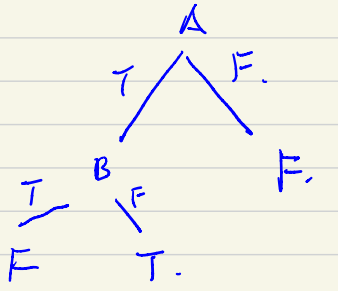
Based on the Information Gain criteria, we would also want to choose X_1 as the first feature to split since it has the largest information Gain.

4.2.

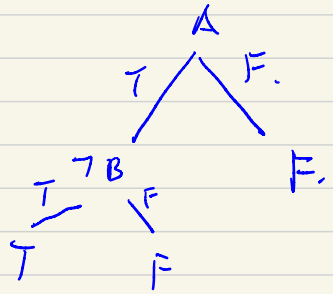
① $A \wedge \neg B$.

A	B	¬B	A and ¬B
T	F	T	T
T	T	F	F
F	F	T	F
F	T	F	F.

Using decision Tree.



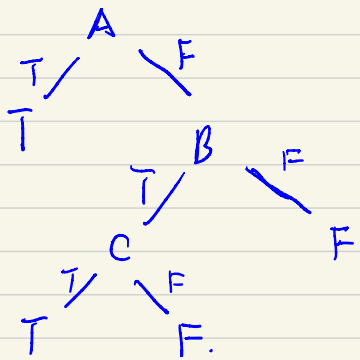
or.



② $A \vee (B \wedge C)$

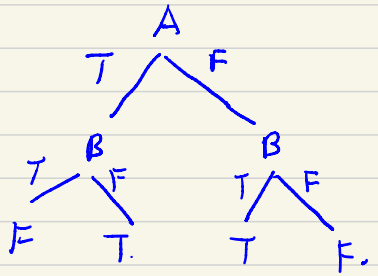
A	B	C	A or (B and C).
F	F	F	F
F	F	T	F
F	T	F	F
F	T	T	T
T	F	F	T
T	F	T	T
T	T	F	T
T	T	T	T.

By decision tree:



(3). A XOR B.

A	B	A XOR B.
F	F	F
F	T	T
T	F	T
T	T	F.



(4). (A ∧ B) ∨ (C ∧ D).

A	B	C	D	(A∧B)∨(C∧D).
F	F	F	F	F
F	F	F	T	F
F	F	T	F	F
F	F	T	T	T
F	T	F	F	F
F	T	F	T	F
F	T	T	F	T
F	T	T	T	T
T	F	F	F	F
T	F	F	T	F
T	F	T	F	T
T	F	T	T	T
T	T	F	F	T
T	T	F	T	T
T	T	T	F	T
T	T	T	T	T