Homework 1
Machine Learning

1. Correlation

Pearson correlation on Spearman on each $x_i$ and $x_j$

$x_1 = (1, 2, 3, 4, 5)$
$x_2 = (10, 8, 6, 4, 2)$

| $x_1$ | $x_2$ |
|---|---|
| 1 | 10 |
| 2 | 8 |
| 3 | 6 |
| 4 | 4 |
| 5 | 2 |

Pearson: $\frac{\mathrm{Cov}(x_1, x_2)}{\sqrt{\mathrm{Var}(x_1)}\sqrt{\mathrm{Var}(x_2)}}$

$\mathrm{Cov}(x_1, x_2) = \frac{1}{n-1}\sum_{i=1}^{n}(x_1 - \bar{x}_1)(x_2 - \bar{x}_2)$

$\bar{x}_1 = \frac{1+2+3+4+5}{5} = 3$, $\bar{x}_2 = \frac{10+8+6+4+2}{5} = 6$

$= \frac{1}{4}((1-3)(10-6) + (2-3)(8-6) + (3-3)(6-6) + (4-3)(4-6) + (5-3)(2-6))$

$= \frac{1}{4}(-2 \cdot 4 + -1 \cdot 2 + 0 \cdot 0 + 1 \cdot -2 + 2 \cdot -4)$

$= \frac{1}{4}(-8 - 2 - 2 - 8)$

$= \frac{1}{4}(-20)$

$= -5$

$\sqrt{\mathrm{Var}(x_1)} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_1 - \bar{x}_1)^2}$

$\sqrt{\mathrm{Var}(x_1)} = \sqrt{\frac{1}{4}((1-3)^2 + (2-3)^2 + (3-3)^2 + (4-3)^2 + (5-3)^2)}$

$= \sqrt{\frac{1}{4}(4 + 1 + 0 + 1 + 4)}$

$= \sqrt{\frac{10}{4}}$

$= \sqrt{\frac{5}{2}}$

$\sqrt{\mathrm{Var}(x_2)} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_2 - \bar{x}_2)^2}$

$= \sqrt{\frac{1}{4}((10-6)^2 + (8-6)^2 + (6-6)^2 + (4-6)^2 + (2-6)^2)}$

$= \sqrt{\frac{1}{4}(16 + 4 + 0 + 4 + 16)}$

$= \sqrt{\frac{40}{4}}$

$= \sqrt{10}$

$r = \frac{-5}{\sqrt{\frac{5}{2}}\sqrt{10}} = \frac{-5}{\sqrt{25}} = -1$

Spearman: Rank $x_1$ and $x_2$

| $x_1$ | Rank $x_1$ | $x_2$ | Rank $x_2$ |
|---|---|---|---|
| 1 | 1 | 10 | 1 |
| 2 | 2 | 8 | 2 |
| 3 | 3 | 6 | 3 |
| 4 | 4 | 4 | 4 |
| 5 | 5 | 2 | 5 |

apply Pearson formula

Numerator:

$\sum_{i=1}^{n}(x_1 - \bar{x}_1)(x_2 - \bar{x}_2)$

$= (-2 \cdot 4) + (-1 \cdot 2) + (0 \cdot 0) + (1 \cdot -2) + (2 \cdot -4)$

$= (-8) + (-2) + (0) + (-2) + (-8)$

$= -20$

Denominator:

$\sqrt{\sum_{i=1}^{n}(x_1 - \bar{x}_1)^2}\sqrt{\sum_{i=1}^{n}(x_2 - \bar{x}_2)^2}$

$= \sqrt{(-2)^2 + (-1)^2 + (0)^2 + (1)^2 + (2)^2}\sqrt{(4)^2 + (2)^2 + (0)^2 + (-2)^2 + (-4)^2}$

$= \sqrt{4 + 1 + 0 + 1 + 4}\sqrt{16 + 4 + 0 + 4 + 16}$

$= \sqrt{10}\sqrt{40}$

$= \sqrt{10 \cdot 40}$

$= \sqrt{400}$

$= 20$

$r = \frac{-20}{20} = -1$

Spearman: $\frac{\sum_{i=1}^{n}(x_1 - \bar{x}_1)(x_2 - \bar{x}_2)}{\sqrt{\sum_{i=1}^{n}(x_1 - \bar{x}_1)^2}\sqrt{\sum_{i=1}^{n}(x_2 - \bar{x}_2)^2}}$

$= \frac{-20}{20} = -1$

$x_1$ and $x_2$ are perfectly linearly correlated and perfectly rank correlated

Since the formula for Pearson correlation and Spearman correlation are identical, the addition of $x_3$ will be, in computation and likelihood

b) Pearson (Logistic function)

$x_3 = f(x_2) = \frac{1}{1+e^{x_2}}$

| $x_2$ | $x_3$ |
|---|---|
| -4 | 0.016 |
| -3 | 0.119 |
| -2 | 0.119 |
| -1 | 0.119 |
| 0 | 0.119 |
| 1 | 0.016 |
| 2 | 0.016 |

$\bar{x}_1 = 0$
$\bar{x}_2 = 0.5$

Numerator:

$\sum_{i=1}^{n}(x_1 - \bar{x}_1)(x_2 - \bar{x}_2)$

$= (-2 \cdot 0.016) + (-1 \cdot 0.119) + (0 \cdot 0.119) + (1 \cdot 0.119) + (2 \cdot 0.016)$

$= (-0.032) + (-0.119) + (0) + (0.119) + (0.032)$

$= 0$

Denominator:

$\sqrt{\sum_{i=1}^{n}(x_1 - \bar{x}_1)^2}\sqrt{\sum_{i=1}^{n}(x_2 - \bar{x}_2)^2}$

$= \sqrt{(-2)^2 + (-1)^2 + (0)^2 + (1)^2 + (2)^2}\sqrt{(-0.016)^2 + (-0.119)^2 + (0.119)^2 + (0.119)^2 + (0.016)^2}$

$= \sqrt{4 + 1 + 0 + 1 + 4}\sqrt{0.016 + 0.014 + 0.014 + 0.014 + 0.016}$

$= \sqrt{10}\sqrt{0.074}$

$= \sqrt{10 \cdot 0.074}$

$= \sqrt{0.74}$

$= 0.859$

$r = \frac{0}{0.859} = 0$

Spearman:

| $x_2$ | Rank $x_2$ | $x_3$ | Rank $x_3$ |
|---|---|---|---|
| -4 | 1 | 0.016 | 1 |
| -3 | 2 | 0.119 | 2 |
| -2 | 3 | 0.119 | 3 |
| -1 | 4 | 0.119 | 4 |
| 0 | 5 | 0.119 | 5 |
| 1 | 6 | 0.016 | 6 |
| 2 | 7 | 0.016 | 7 |

Between Rank $x_2$ and Rank $x_3$ and $x_2$ and $x_3$ are the same as exercise a), so Spearman coefficient is -1.

$r_{\text{Spearman}} = -1$

$x_3$ and $x_2$ are linearly correlated and rank correlated

As the ranks are exactly the same, the same conclusion will be 1

The $x_3$ is also quite high indicating that in addition to linear linear

2. ID3 algorithm → decision tree

Which feature first separates the classes

Wait: $\frac{6}{12}$, $\frac{6}{12}$ Entropy

$E_{\text{Wait}} = -\left(\frac{6}{12}\log_2\frac{6}{12} + \frac{6}{12}\log_2\frac{6}{12}\right)$

$= -\left(\frac{1}{2}\log_2\frac{1}{2} + \frac{1}{2}\log_2\frac{1}{2}\right)$

$= 1.0$

Weather

Sunny: $\frac{2}{4}$, $\frac{2}{4}$
Rain: $\frac{2}{4}$, $\frac{2}{4}$
Cloudy: $\frac{2}{4}$, $\frac{2}{4}$

Sunny: Wait Yes? Wait No?
Rain: Wait Yes? Wait No?
Cloudy: Wait Yes? Wait No?

Sunny: $\frac{2}{4}$ Wait Yes, $\frac{2}{4}$ Wait No

$E_{\text{Sunny}} = -\left(\frac{2}{4}\log_2\frac{2}{4} + \frac{2}{4}\log_2\frac{2}{4}\right)$

$= -\left(\frac{1}{2}\log_2\frac{1}{2} + \frac{1}{2}\log_2\frac{1}{2}\right)$

$= 1.0$

Rain: $\frac{2}{4}$ Wait Yes, $\frac{2}{4}$ Wait No

$E_{\text{Rain}} = -\left(\frac{2}{4}\log_2\frac{2}{4} + \frac{2}{4}\log_2\frac{2}{4}\right)$

$= -\left(\frac{1}{2}\log_2\frac{1}{2} + \frac{1}{2}\log_2\frac{1}{2}\right)$

$= 1.0$

Cloudy: $\frac{2}{4}$ Wait Yes, $\frac{2}{4}$ Wait No

$E_{\text{Cloudy}} = -\left(\frac{2}{4}\log_2\frac{2}{4} + \frac{2}{4}\log_2\frac{2}{4}\right)$

$= -\left(\frac{1}{2}\log_2\frac{1}{2} + \frac{1}{2}\log_2\frac{1}{2}\right)$

$= 1.0$

$IG = 1.0 - 1.0 = 0.0$

Weather → Sunny: W. Yes, W. No Cloudy

$E_{\text{Sunny}} = -\left(\frac{1}{2}\log_2\frac{1}{2} + \frac{1}{2}\log_2\frac{1}{2}\right)$

$= -\left(\frac{1}{2}\log_2\frac{1}{2} + \frac{1}{2}\log_2\frac{1}{2}\right)$

$= 1.0$

$E_{\text{Rain}} = -\left(\frac{1}{2}\log_2\frac{1}{2} + \frac{1}{2}\log_2\frac{1}{2}\right)$

$= -\left(\frac{1}{2}\log_2\frac{1}{2} + \frac{1}{2}\log_2\frac{1}{2}\right)$

$= 1.0$

$E_{\text{Cloudy}} = -\left(\frac{1}{2}\log_2\frac{1}{2} + \frac{1}{2}\log_2\frac{1}{2}\right)$

$= -\left(\frac{1}{2}\log_2\frac{1}{2} + \frac{1}{2}\log_2\frac{1}{2}\right)$

$= 1.0$

$IG = 1.0 - 1.0 = 0.0$

Wind: $\frac{3}{4}$ Yes, $\frac{1}{4}$ No

$E_{\text{Wind}} = -\left(\frac{3}{4}\log_2\frac{3}{4} + \frac{1}{4}\log_2\frac{1}{4}\right)$

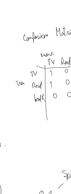
$= -\left(\frac{3}{4}\log_2\frac{3}{4} + \frac{1}{4}\log_2\frac{1}{4}\right)$

$= 0.811$

$IG = 1.0 - 0.811 = 0.189$

Weather has the largest IG, therefore it should be the root of the decision tree

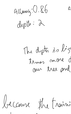
b)




Confusion Matrix

| | Wait Yes | Wait No |
|---|---|---|
| TP | 1 | 0 |
| FP | 0 | 2 |
| FN | 0 | 0 |

3.




The depth has bigger impact on the tree because more data for the training of the tree and with 100 vs 1000

b) Because the training set and the test set have the same proportion of examples of each class when stratifying is applied. Therefore, without stratifying, the accuracy lowers.