Assignment 2

Date :

$$P(Y=y|X_1,X_2) = P(y=1|X_1,X_2)^y \, P(y=0|X_1,X_2)^{1-y}$$

$$= \left[\frac{1}{1+\exp(-(\beta_0+\beta_1X_1+\beta_2X_2))}\right]^y \left[\frac{\exp(-(\beta_0+\beta_1X_1+\beta_2X_2))}{1+\exp(-(\beta_0+\beta_1X_1+\beta_2X_2))}\right]^{1-y}$$

$$= \frac{\exp(-(\beta_0+\beta_1X_1+\beta_2X_2)(1-y))}{1+\exp(-(\beta_0+\beta_1X_1+\beta_2X_2))}$$

$$L(\beta) = \prod_{i=1}^{n_1+n_2} P(Y=y_i|X_{1_i},X_{2_i})$$

$$= \prod_{i=1}^{n_1+n_2} \frac{\exp(-(\beta_0+\beta_1X_{1_i}+\beta_2X_{2_i})(1-y_i))}{1+\exp(-(\beta_0+\beta_1X_{1_i}+\beta_2X_{2_i}))}$$

$$= \frac{\exp(-\sum_{i=1}^{n}(\beta_0+\beta_1X_{1_i}+\beta_2X_{2_i})(1-y_i))}{\prod_{i=1}^{n_1+n_2}(1+\exp(-(\beta_0+\beta_1X_{1_i}+\beta_2X_{2_i})))}$$

$$LL(\beta) = -\sum_{i=1}^{n_1+n_2}(\beta_0+\beta_1X_{1_i}+\beta_2X_{2_i})(1-y_i) - \sum_{i=1}^{n_1+n_2}\log\left(\frac{1+\exp(\beta_0+\beta_1X_{1_i}+\beta_2X_{2_i})}{\exp(\beta_0+\beta_1X_{1_i}+\beta_2X_{2_i})}\right)$$

$$= \sum_{i=1}^{n_1+n_2}\left(y_i(\beta_0+\beta_1X_{1_i}+\beta_2X_{2_i}) - \log(1+\exp(\beta_0+\beta_1X_{1_i}+\beta_2X_{2_i}))\right)$$

$$P(Y=1|X_1,X_2) = \frac{1}{1+e^{-(\beta_0+\beta_1X_1+\beta_2X_2)}}$$

$$= \frac{1}{1+e^{-(-6+0.05\times40+3.5)}} = \frac{1}{1+e^{0.5}} = \frac{1}{1+\sqrt{e}} \approx 0.3775$$

for 50%, $e^{-(\beta_0+\beta_1X_1+\beta_2X_2)} = 1 \Rightarrow \beta_1X_1+\beta_2X_2+\beta_0 = 0$

$$\Rightarrow X_1 = \frac{-\beta_0-\beta_2X_2}{\beta_1}$$

$$= \frac{6-3.5}{0.05} = 50$$

$\therefore$ 50 hrs is needed

$$\cancel{P(div} \; P(\text{dividend} \mid X) = \frac{f(X \mid \text{dividend})\, P(\text{dividend})}{f(X \mid \text{dividend}) P(D) + P(nD) f(X \mid nD)}$$

$$P(\text{dividend} / X) = \frac{f(X \mid \text{dividend } D)\, P(D)}{f(X \mid D) P(D) + P(nD) f(X \mid nD)}$$

$$= \frac{1}{1 + \frac{0.2}{0.8} \cdot \exp\left(\frac{-1}{72}\left((0-4)^2 - (10-4)^2\right)\right)}$$ [Assuming gaussian]

$$= \frac{1}{1 + \frac{1}{4} e^{5/18}}$$

$$\approx 0.752$$

$$\frac{jC_1(C_j C_2 + R_2)}{j(C_1 + C_2) + R_2}$$

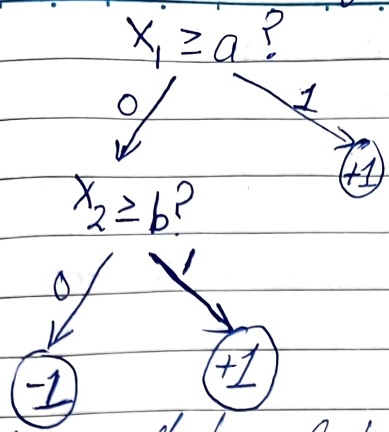
I_1 $\quad$ I_2

$I_1 C_1 = \quad C_1 \frac{dV_1}{\ldots}$

Assignment 3

Date :

Q1.



Q2. (a) It's a collection of trees... decision trees.

(b) It was built by bootstrapping, which involves randomly selecting data, and more importantly, each tree is built independently by selecting a random subset of features to train each node on. ("train" i.e. choose which feature from the subset gives the least gini impurity on the training dataset).

(c) It does not 'help improve performance', it is simply for a different task. Decision trees deterministically fit the ~~data~~ training data and hence are not used for classifying new data. By introducing randomness and majority into the training process, we avoid overfitting the training data and get ~~to~~ flexibility to ~~test~~ classify new data more accurately. The model is forced to find patterns in the training data rather than brute-force classify a single known set of data.

Q3. (a) ~~By~~ Ensemble methods are strategies where you aggregate the result of multiple trained models and use that result as the prediction.

(b) Models have bias, by aggregating the results of multiple models you are averaging their biases, hence resulting in a lower bias final model due to cancelling of biases.

(c) Yes. By definition.

Q4. 1. TP = 180 FN = 20
FP = 70 TN = 730

2. Accuracy = $\frac{\text{correct}}{\text{total}} = \frac{910}{1000} = 0.91$

~~Precision~~ Recall = $\frac{\text{correct pos}}{\text{actual total pos}} = \frac{180}{200} = 0.9$

Precision ~~Recall~~ = $\frac{\text{Correct pos}}{\text{predicted total pos}} = \frac{180}{250} = 0.72$

Specificity = $\frac{\text{Correct neg}}{\text{actual total neg}} = \frac{730}{800} = \cancel{0.912}$ 0.9125

Date :

$$F1\ Score = \frac{2 \cdot Pr \cdot R}{Pr + R} = \frac{2 \times 0.72 \times 0.9}{0.9 + 0.72} = 0.8$$

Recall. $Recall = \frac{1}{1 + \frac{FN}{TP}}$, thus it's low for large false negatives

If you lower the threshold, more positives will be ~~det~~ predicted, hence both TP and FP will increase.

Accuracy only depends on the sum of ~~corre~~ true predictions TP+TN and sum of false predictions FP+FN. The actual confusion matrix may vary drastically, with all the sum contribution coming from one term (say TP) vs an equal distribution (50% TP 50% TN)