CS540 Introduction to Artificial Intelligence

Lecture 6

Young Wu

Based on lecture slides by Jerry Zhu, Yingyu Liang, and Charles Dyer

July 5, 2022

Hat Game

Quiz

Hat Game Diagram

Discussion

Axes Aligned Decision Boundary

Motivation

Decision Tree

Description

- Find the feature that is the most informative.
- Split the training set into subsets according to this feature.
- Repeat on the subsets until all the labels in the subset are the same.

Binary Entropy

Definition

- Entropy is the measure of uncertainty.
- The value of something uncertain is more informative than the value of something certain.
- For binary labels, $y_i \in \{0, 1\}$, suppose p_0 fraction of labels are 0 and $1 - p_0 = p_1$ fraction of the training set labels are 1, the entropy is:

$$H(Y) = p_0 \log_2 \left(\frac{1}{p_0} \right) + p_1 \log_2 \left(\frac{1}{p_1} \right)$$
$$= -p_0 \log_2 (p_0) - p_1 \log_2 (p_1)$$

Entropy

Definition

- If there are K classes and p_y fraction of the training set labels are in class y, with $y \in \{1, 2, ..., K\}$, the entropy is:

$$H(Y) = \sum_{y=1}^{K} p_y \log_2 \left(\frac{1}{p_y}\right)$$
$$= -\sum_{y=1}^{K} p_y \log_2 (p_y)$$

Entropy

Quiz

Entropy Math

Quiz

Entropy 2

Quiz

Conditional Entropy

Definition

- Conditional entropy is the entropy of the conditional distribution. Let K_X be the possible values of a feature X and K_Y be the possible labels Y. Define p_x as the fraction of the instances that are x, and $p_{y|x}$ as the fraction of the labels that are y among the ones with instance x.

$$H\left(Y|X = x\right) = -\sum_{y=1}^{K_Y} p_{y|x} \log_2\left(p_{y|x}\right)$$

$$H\left(Y|X\right) = \sum_{x=1}^{K_X} p_x H\left(Y|X = x\right)$$

Aside: Cross Entropy

Definition

- Cross entropy measures the difference between two distributions.

$$H(Y, X) = -\sum_{z=1}^{K} p_{Y=z} \log_2 \left(p_{X=z}\right)$$

- It is used in logistic regression to measure the difference between actual label Y_i and the predicted label A_i for instance i, and at the same time, to make the cost convex.

$$H(Y_i, A_i) = -y_i \log(a_i) - (1 - y_i) \log(1 - a_i)$$

Information Gain

Definition

- The information gain is defined as the difference between the entropy and the conditional entropy.

$$I(Y|X) = H(Y) - H(Y|X).$$

- The larger than information gain, the larger the reduction in uncertainty, and the better predictor the feature is.

Splitting Discrete Features

Definition

- The most informative feature is the one with the largest information gain.

$$\underset{j}{\operatorname{argmax}}\, I\left(Y|X_j\right)$$

- Splitting means dividing the training set into K_{X_j} subsets.

$$\left\{(x_i, y_i) : x_{ij} = 1\right\}, \left\{(x_i, y_i) : x_{ij} = 2\right\}, ..., \left\{(x_i, y_i) : x_{ij} = K_{X_j}\right\}$$

Splitting Continuous Variables Diagram

Definition

ID3 Algorithm (Iterative Dichotomiser 3)

Description

Pruning Diagram

Discussion

Bagging Diagram

Discussion

Boosting Diagram

Discussion

K Nearest Neighbor

Description

- Given a new instance, find the K instances in the training set that are the closest.
- Predict the label of the new instance by the majority of the labels of the K instances.

Distance Function

Definition

- Many distance functions can be used in place of the Euclidean distance.

$$\rho\left(x, x'\right) = \left\|x - x'\right\|_2 = \sqrt{\sum_{j=1}^{m}\left(x_j - x'_j\right)^2}$$

- An example is Manhattan distance.

$$\rho\left(x, x'\right) = \sum_{j=1}^{m}\left|x_j - x'_j\right|$$

Manhattan Distance Diagram

Definition

1 Nearest Neighbor

Quiz

3 Nearest Neighbor

Quiz

K Fold Cross Validation

Discussion

5 Fold Cross Validation Example

Discussion

Leave One Out Cross Validation

Discussion

Cross Validation

Quiz

Cross Validation 2

Quiz

Lecture Next Week

Admin