HW2

ID	Name
1953902	GAO Yangfan

Table of Contents

Problem 1

Explanation

Supervised learning is a set of learning methods that requires the training data labeled, which is mostly used in regression and classification while unsupervised learning methods don't require labeled training data, which is mostly used in clustering and data dimension reduction.

Algorithms

For supervised learning: linear regression, softmax regression

For unsupervised learning: PCA, K-means.

Problem 2

The MSE is defined as:

$$\frac{1}{m}\sum_{i=1}^{m}(\hat{y}-y)^2$$

where m is the size of sample set, $\hat{y}$ is the predicted value and y is the ground truth.

In this specified problem, we have $m=3$ and there are 3 situations:

1. Leave $(0,2)$ as the test example, the fitted parameters by the rest data are:

$$w=\frac{1-2}{3-2}=-1$$
$$b=2-w*2=4$$

The predicted value of the test example is:

$$\hat{y}=w*0+b=4$$

The MSE of this situation is:

$$(4-2)^2=4$$

Similarly, we can get the MSE of other situations:

2. Leave $(2,2)$ as the test example, the MSE is

$$(\frac{4}{3}-2)^2=\frac{4}{9}$$

3. Leave $(3,1)$ as the test example, the MSE is

$$(2-1)^2=1$$

The average MSE is:

$$\frac{4 + \dfrac{4}{9} + 1}{3} = \frac{49}{27}$$

Problem 3

1.

$$\nabla u = [\frac{\partial u}{\partial x}, \frac{\partial u}{\partial y}]^T = [y, x + 2y]^T$$

2.

$$\nabla u = [\frac{\partial u}{\partial x}, \frac{\partial u}{\partial y}, \frac{\partial u}{\partial z}]^T = [\frac{x}{x^2 + y^2 + z^2}, \frac{y}{x^2 + y^2 + z^2}, \frac{z}{x^2 + y^2 + z^2}]^T$$
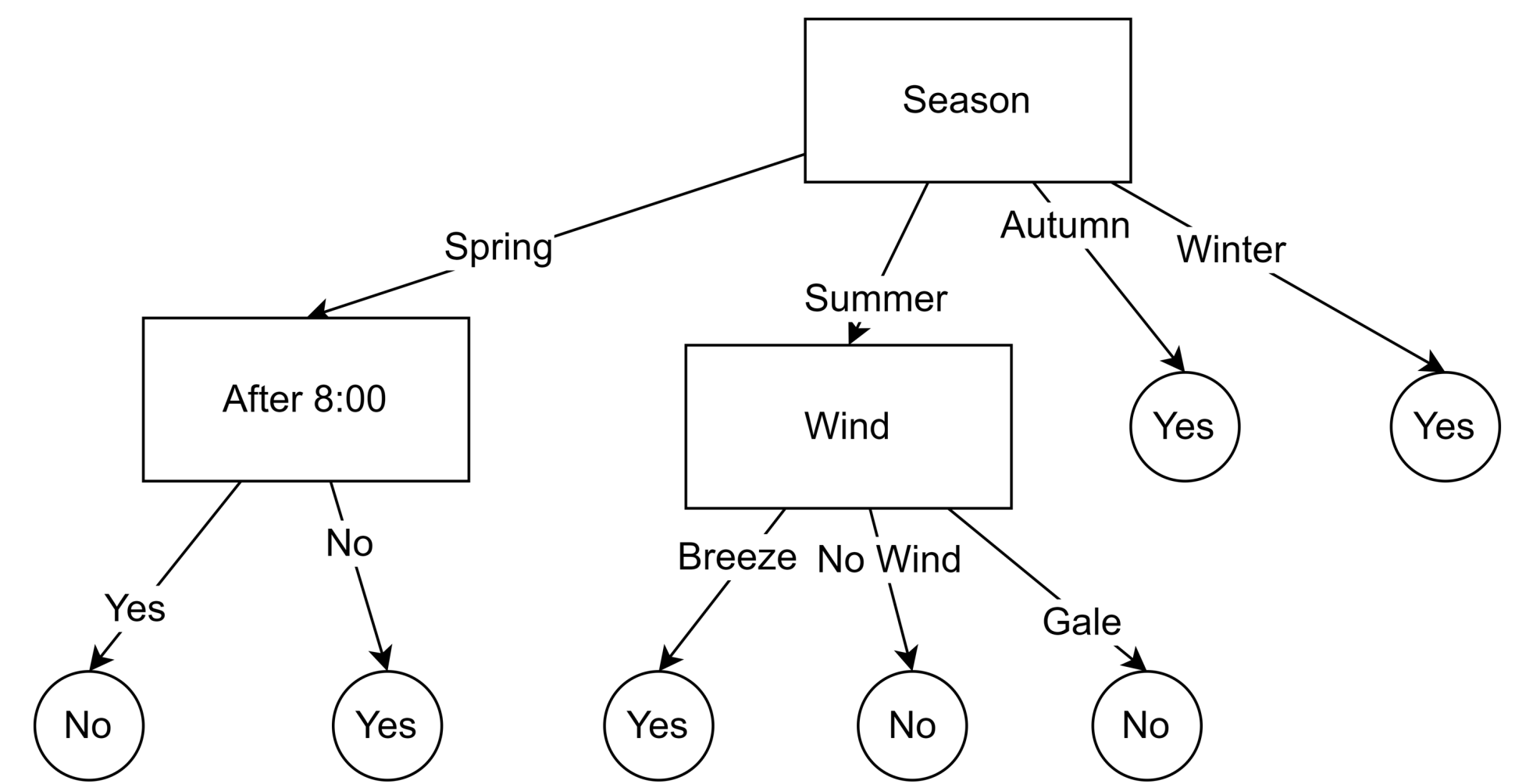
plug (1,2,-2)$ into the result, we get:

$$\nabla u|_{(1,2,-2)} = [\frac{1}{9}, \frac{2}{9}, -\frac{2}{9}]^T$$

Problem 4

I will first display the result and then illustrate the ideas when generating the decision tree.

   1. result

Note: Rectangle nodes are internal nodes while the circle ones are leaf nodes

   2. Idea illustration

Actually, to solve this problem, we don't need to compute all of the conditional entropy of each situation.

For the root node, it is easy to discover that if the season is autumn or winter, student A will always sleep in by observation, while other features don't have this property. So using 'season' as the root node implies maximum information gain.

As autumn and winter is determine to be 'sleep in', we then shuold just take spring and summer into consideration.

For 'spring' branch, there are just 2 examples, so no matter what feature we choose, the information gain will be maximized. But for feature 'wind', we don't have a value of 'gale' on this subset and the number of positive and negative examples is equal, so it is hard to determine which class it shall be for 'gale', so I chose the feature 'After 8:00'

For 'summer' branch, we have to compute the conditional entropy. We can easily compute that if we choose 'wind' as the feature, the entropy will be 0, which implies the maximum information gain. As there are no data that take the value 'no wind', the leaf node is marked with the class that appears the most time in this branch, which is 'not to sleep in'.

Problem 5

According to the assumptions and steps of Naive Bayes, we should compute

$$P(y = 1) \prod_{i=1}^{2} P(x_i|y = 1)$$

$$P(y = -1) \prod_{i=1}^{2} P(x_i|y = -1)$$

and choose the bigger value as the final class

According to the given dataset, we can compute:

$$P(y = 1) \prod_{i=1}^{2} P(x_i|y = 1) = P(y = 1) * P(x_1 = 2|y = 1) * P(x_2 = S|y = 1) = \frac{2}{3} * \frac{2}{5} * \frac{1}{10} = \frac{2}{75}$$

$$P(y = -1) \prod_{i=1}^{2} P(x_i|y = -1) = P(y = -1) * P(x_1 = 2|y = -1) * P(x_2 = S|y = -1) = \frac{1}{3} * \frac{1}{5} * \frac{3}{5} = \frac{3}{75}$$

as $\frac{3}{75} > \frac{2}{75}$, the final class for the given data $(2, S)$ is $y = -1$