Student ID:

Name:

Question 1: Consider the following data set:

Restaurant	Type	Price	Neighborhood	Restriction	OK
R_1	Fast Food	$	Oakland	Vegetarian	0
R_2	Ethnic	$$	Squirrel Hill	Gluten Free	0
R_3	Casual Dining	$$	Squirrel Hill	None	0
R_4	Casual Dining	$$$	Shadyside	Vegetarian	0
R_5	Casual Dining	$	Oakland	Vegetarian	1
R_6	Fast Food	$$	Squirrel Hill	None	1
R_7	Ethnic	$	Squirrel Hill	None	1
R_8	Casual Dining	$	Shadyside	Gluten Free	0
R_9	Fast Food	$$$	Oakland	None	0
R_{10}	Ethnic	$$	Shadyside	Vegetarian	1
R_{11}	Casual Dining	$$	Shadyside	Gluten Free	1

Suppose we decide to construct a decision tree using multiple splits and the Gini index impurity measure. Which attributes would be the best to use as the root node assuming that we consider each of the input features to be unordered? (12 points)

Solutions:

Let's look at the first level. We will be trying all attributes and checking which one yields the maximum information gain upon splitting.

Information gain of the whole dataset is:

$1 - \frac{6}{11} \times \frac{6}{11} - \frac{5}{11} \times \frac{5}{11} = \frac{60}{121}$

- Type: when type is used as an attribute, we get the following split:

Value	OK=0	OK=1	Gini
Fast food	2	1	$1-(\frac{2}{3})^2-(\frac{1}{3})^2 = \frac{4}{9} = 0.44$
Casual dining	3	2	$1-(\frac{3}{5})^2-(\frac{2}{5})^2 = \frac{12}{25} = 0.48$
Ethnic	1	2	$1-(\frac{1}{3})^2-(\frac{2}{3})^2 = \frac{4}{9} = 0.44$

Conditional gini turns out to be: $\frac{3}{11} \times \frac{4}{9} + \frac{5}{11} \times \frac{12}{25} + \frac{3}{11} \times \frac{4}{9} = 0.46$

- Price: when price is used as an attribute, we get the following split:

Value	OK=0	OK=1	Gini
$	2	2	0.5
$$	2	3	0.48
$$$	2	0	0

Conditional gini turns out to be: $\frac{4}{11} \times 0.5 + \frac{5}{11} \times 0.48 + 0 = 0.4$

- Neighborhood: when neighborhood is used as an attribute, we get:

Value	OK=0	OK=1	Gini
Oakland	2	1	0.44
Squirrel Hill	2	2	0.5
Shadyside	2	2	0.5

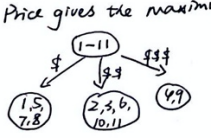
Conditional gini turns out to be: $\frac{3}{11} \times 0.44 + \frac{4}{11} \times 0.5 + \frac{4}{11} \times 0.5 = 0.48$

- Restriction: When restriction is used as an attribute, we get:

Value	OK=0	OK=1	Gini
Vegetarian	2	2	0.5
Gluten Free	2	1	0.44
None	2	2	0.5

Conditional gini turns out to be: $\frac{3}{11} \times 0.44 + \frac{8}{11} \times 0.5 = 0.48$

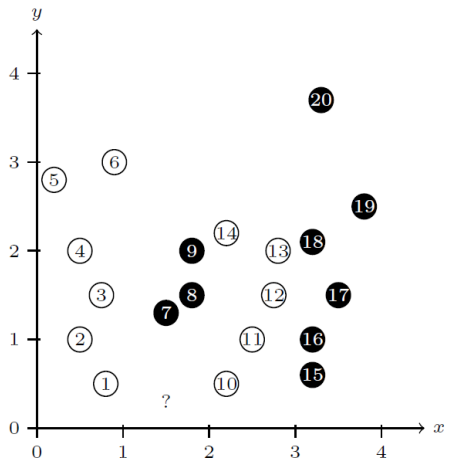
Splitting on Price gives the maximum information gain. The decision tree is:



Price is the best attribut to use as the root node.

Question 2: For the given datasets with the k-Nearest Neighbors algorithm. As shown in the following figure, we show a set of training points classified as being either black or white. Please illustrate the reasons to get the full marks.

1 Predict the point marked by "?" with k=1,2,3 and Euclidean distance (2 points)
2 Are there any points in the training set that would be misclassified using k=1 using Euclidean distance? If so, identify them. (2 points)
3 If we modify the distance metric to 'the distance on the x-axis', please answer the question 1 and 2 again. (4 points)
4 What happens when k=5 using your distance metric? (2 points)



Solution:

1. It's classified as white for k = 1, 2, and 3.
2. Yes, points 9, 14, 13, and 18 would be misclassified using k = 1. (Of course we mean if that point wasn't in the training dataset it would be misclassified by k = 1.)
3. 1) classified as black if k=1,2,3
   2) all of them will be right classified.

4. Using k = 5 some points in the training set would be misclassified: points 7, 8, and 9 would all be classified as white, and points 12 and 13 would be classified as black.

Question 3: Consider the training data set. There are three attributes A, B, and C. The class label is in column Y.
(1) Predict the class label for a test sample (A=1, B=0, C=0) using the naïve Bayes classifier. The answer can be +, -, or cannot decide. (10 points)
(2)We modify the data record 9 from (A: 0, B: 2, C=0) to (A: 0, B: 0, C=1). Then predict the class label for a test sample (A=1, B=0, C=0) using the naïve Bayes classifier. (10 points)

Record	A	B	C	Y
1	1	0	1	+
2	0	2	0	-
3	1	1	0	-
4	0	1	1	+

5	0	0	0	+
6	0	2	1	-
7	1	1	0	+
8	1	2	1	+
9	0	2	0	-
10	1	1	1	+

Solution:

(1) $P(Y=+|A=1,B=0,C=0) = P(A=1|+)P(B=0|+)P(C=0|+)P(+)/P(1,0,0)$

$= \frac{4}{6} \times \frac{2}{6} \times \frac{2}{6} \times 0.6 / P(1,0,0) = \frac{0.044}{P(1,0,0)}$

$P(Y=-|A=1,B=0,C=0) = P(A=1|-)P(B=0|-)P(C=0|-)P(-)/P(1,0,0)$

$= \frac{1}{4} \times 0 \times \frac{3}{4} \times 0.4 / P(1,0,0) = \frac{0}{P(1,0,0)}$

So the test sample should be classified into '+'.

(2) $P(Y=+|A=1,B=0,C=0) = P(A=1|+)P(B=0|+)P(C=0|+)P(+)/P(1,0,0)$

$= \frac{4}{6} \times \frac{2}{6} \times \frac{2}{6} \times 0.6 / P(1,0,0) = \frac{0.044}{P(1,0,0)}$

$P(Y=-|A=1,B=0,C=0) = P(A=1|-)P(B=0|-)P(C=0|-)P(-)/P(1,0,0)$

$= \frac{1}{4} \times \frac{1}{4} \times \frac{2}{4} \times 0.4 / P(1,0,0) = \frac{0.0125}{P(1,0,0)}$

So the test sample should be classified into '+'.

Question 4: Check the following binary classifiers that are able to correctly separate the training data (circles vs. triangles) given in following Figure. (8 points)

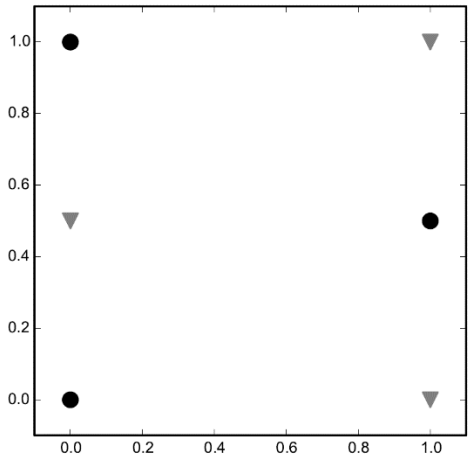
Logistic regression

SVM with linear kernel

Decision tree

3-nearest-neighbor classifier (with Euclidean distance).

Perceptron



Solutions:

- Logistic regression and linear SVM: linear decision functions, hence no.
- 3-NN: the 3 nearest neighbors of any point in our training set are 1 of the same class and 2 of the opposite class, hence 3-NN will be systematically wrong.
- DT: yes, you can partition the space with lines orthogonal to the axes in such away that every sample ends up in a different region.
- Single layer Perceptron is not while multiple layer Perceptron is OK