Discussion 1
Machine Learning, Spring 2019

1 Concepts of learning

Please match each of the following situations with the Classification, Regression, Ranking, Clustering, Conditional probability estimation, Density estimation, Finding patterns problems.

1. Suppose you are working on weather prediction, and your weather station makes one of three predictions for each day's weather: Sunny, Cloudy or Rainy.
2. Suppose you are working at Zillow, and you want to improve house price prediction based on house area, lot size, number of bedrooms, and renovations made.
3. Given 100 images painted by a child, and 100 images painted by an adult, learn to predict the age-group of a new image author (when the identity of this author is unknown).
4. Assume that you just released your first music album and you want to predict the number of copies that will be sold next month.
5. Suppose that you are running an online store and want to interest buyers in products will be the best match for their shopping profile.
6. Market researchers want to partition the general population of consumers into market segments in order to better understand the relationship between different groups of consumers.
7. Assume that you just started as a data scientist at Costco and by analyzing customer baskets you are trying to determine which pair of items are most frequently bought together.
8. A group of people who were at least 21 years old was tested for diabetes mellitus according to World Health Organization criteria. We want to know the skewness and multimodality of this dataset.
9. We might estimate that there is a 30% chance that it will rain tomorrow. Now suppose we have some new information such as looking out the window and noticing that it's cloudy. We want to revise our estimate in light of this new information.
10. Suppose you are working on stock market prediction. You would like to predict whether the Amazon stocks value will go up or down tomorrow.

2 Linear Classifiers: Concepts

1. You are given the following labeled data points: Positive examples: $[-1, 1]$ and $[1, -1]$, Negative examples: $[1, 1]$ and $[2, 2]$. See Figure 1.

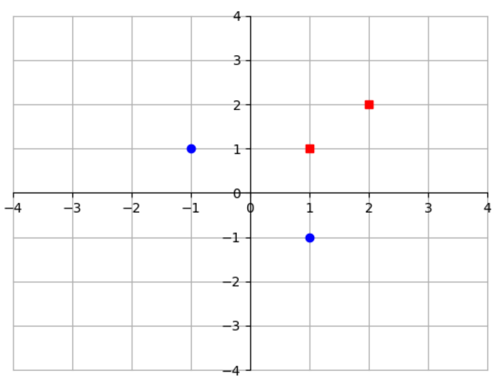
For each of the following parameterized families of classifiers, find the parameters of a classification model in each family that can correctly classify the above data, or explain (with a clear diagram and/or words) why no such choice of parameters exists.

(a) Inside or outside of an origin-centered circle with radius r,

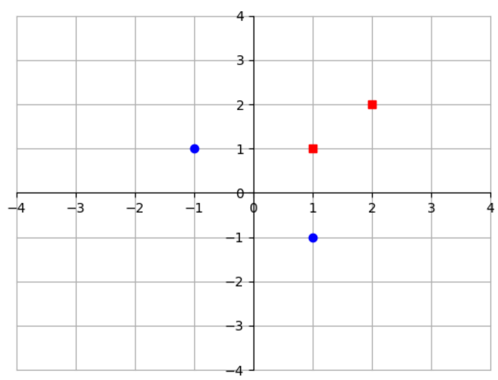
(b) Inside or outside of an $[x, y]$-centered circle with radius r,

(c) Above or below a line through the origin defined by θ,

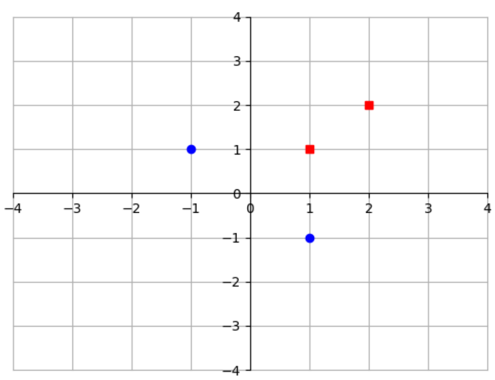
(d) Above or below a line defined by θ and offset θ_0.



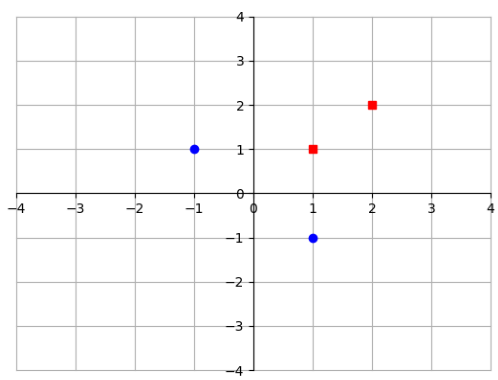
Circle



Circle with offset



Line



Line with offset

Figure 1: 1.1 Circle and Line Classifiers.

2. Which of the above are families of linear classifiers?

3. Suppose we are given a linear classification model that predicts whether or not it is going to rain based upon the temperature (in degrees Celsius) and humidity (expressed as a percentage from 0-100). The model has weights defined such that if the sum of the temperature and the humidity exceeds 110, then it predicts rainfall instead of clear weather.

(a) Assume that an output of $+1$ corresponds to predicted rainfall. This model has a weight vector θ of length 2 and a nonzero offset θ_0. What are the values of θ and θ_0?

(b) Consider what happens when we feed this model a data point from the planet Mercury (where it never rains) on which the temperature is observed to be $400°$ C with a humidity of zero. What does this model predict will happen on Mercury? What does this say about the generalization ability of this model?

4. Consider following dataset (see Figure 2):

$$X = [-2, -1, 0, 1, 2, 3]$$

$$Y = [1, 1, 1, -1, -1, 1]$$

(a) Is the data (X, Y) separable with a linear classifier? If yes, plot the decision boundary.

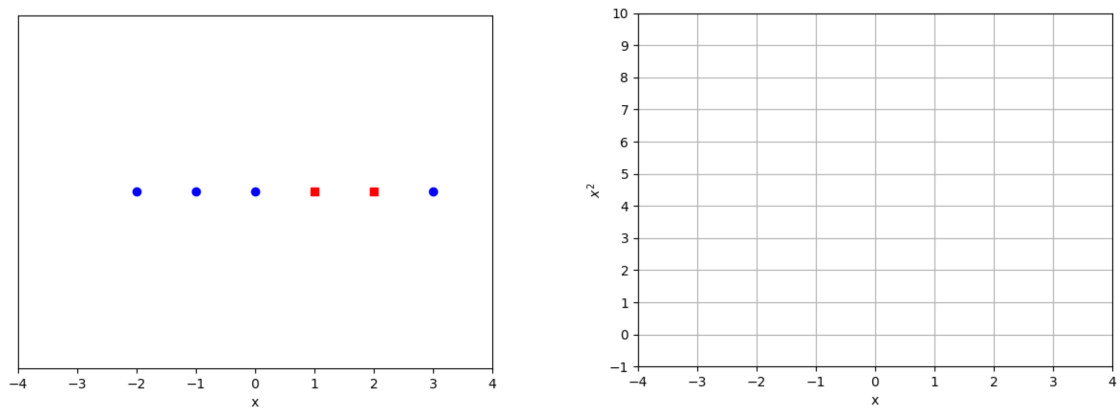
(b) Consider adding new quadratic feature to the data.

$$X_2 = [(-2, 4), (-1, 1), (0, 0), (1, 1), (2, 4), (3, 9)]$$

Plot the training set (X_2, Y) as a labeled scatter in 2 dimensions.

(c) Is the data (X_2, Y) separable with a linear classifier? If yes, plot the decision boundary.

(d) What are advantages and disadvantages of adding higher dimensional features to the dataset?



X $\quad$ X_2

Figure 2: 1.4 Higher dimensions

3 Perceptron

Consider applying the perceptron algorithm (through the origin) on a small training set containing three points: $x_1 = [1, 0], x_2 = [-1, 2], x_3 = [-1, -1]$ with labels $y_1 = 1, y_2 = -1, y_3 = -1$ (see Figure 3). Given that the algorithm starts with $w = [0, 0]$, the first point that the algorithm sees is always considered a mistake. The algorithm starts with one data point and then cycles through the data until it makes no further mistakes.

(a) How many mistakes does the algorithm make until convergence if the algorithm starts with data point x_1? Draw a diagram showing the progression of the plane as the algorithm cycles.

(b) How many mistakes does the algorithm make if it starts with data point x_2? Draw a diagram showing the progression of the plane as the algorithm cycles.

(c) Why there is a difference between the number of mistakes made by the algorithm in part (a) and part (b).

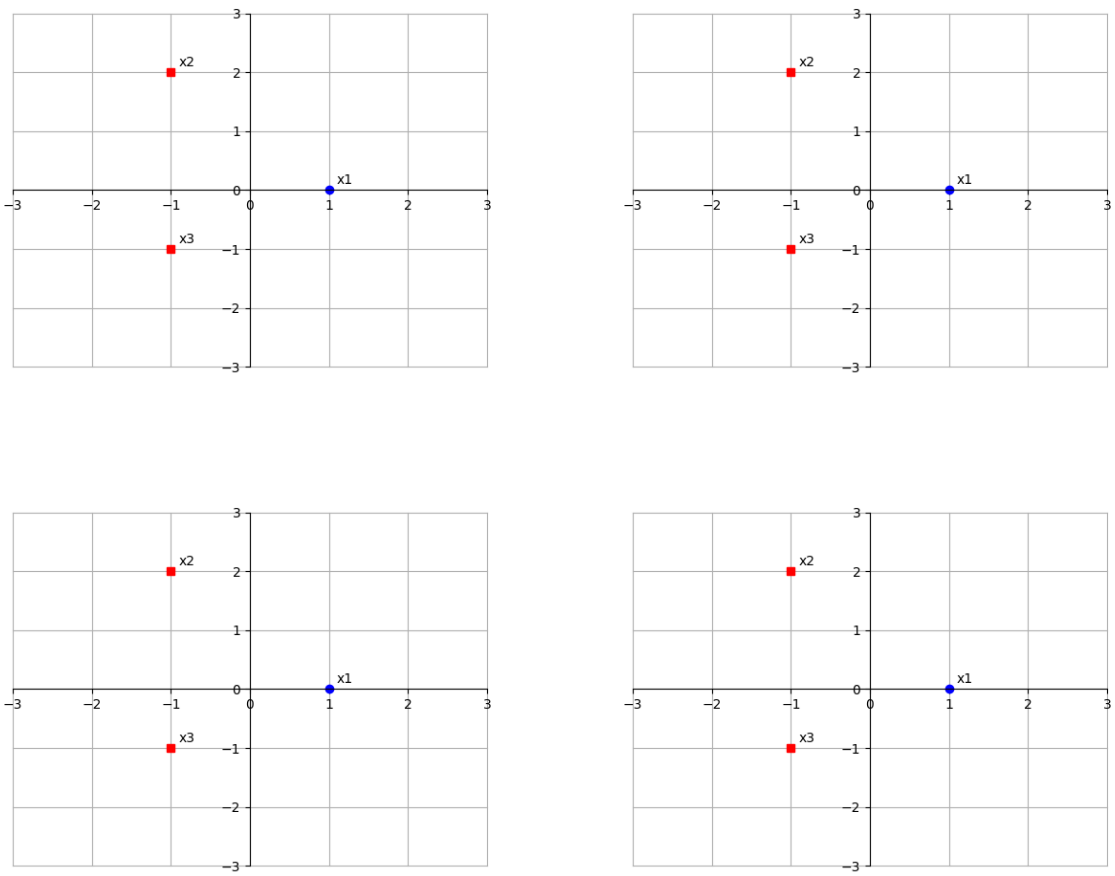


Figure 3: Perceptron.