Data Science and Technology Entrepreneurship

Generative Classifiers, Linear Discriminant Functions VCs and Startups

Sameer Maskey

Week5

Announcements

- Next class
  - 313 FayerWeather (behind Avery)
- Extra Lectures
  - Introduction to Web Programming for Business Students
  - This Friday 3:30 to 5:00
  - Room : TBA
- Google Doc Link Share for Assignment 2
  - Mentors/Advisors are going through Assignment 2
  - Some links are not viewable by mentors

Guest Lecture

- Charlie O Donnell
- Partner, Brooklyn Bridge Ventures

Topics for Today

- Linear Classifiers
- Guest Lecture:
  - What VCs look for in startups?
  - How do you know you have a good idea?
  - When to pivot?

Course Stages

Next few weeks :

Minimum Viable Product Development,
Startup Technology,
Classification Algorithm
Clustering algorithm,
MapReduce,
Customer Validation

Machine Learning and Business

- Methods to analyze data that are all useful in decision making for businesses in general
- Data to Scores
- Data to Classes
  - Discriminative Methods
  - Generative Methods
- Data to Clusters

Data to Classification

- Given a set of features

$$X=(x_1, x_2, x_3, \ldots, x_n)$$

- we want to predict Y

How about x?
How do we get them?

$$Y=\{0,1\}$$

(

, { will not buy OR will buy })

Decision Surface

- We want to find a decision surface that will classify our data better
- Fisher's Linear Discriminant
  - Dimensionality reduction, project data on a line and classify
- Naive Bayes
  - Compute p(y|x) using conditional independence assumption
- Perceptron
  - Linear Discrimination with a hyperplane in (d-1) dimension

Linear Discriminant Functions

- A linear discriminant function is defined by

$$f(x) = w^T x + w_0$$

  - where ‘w’ is the weight vector and w_0 is bias

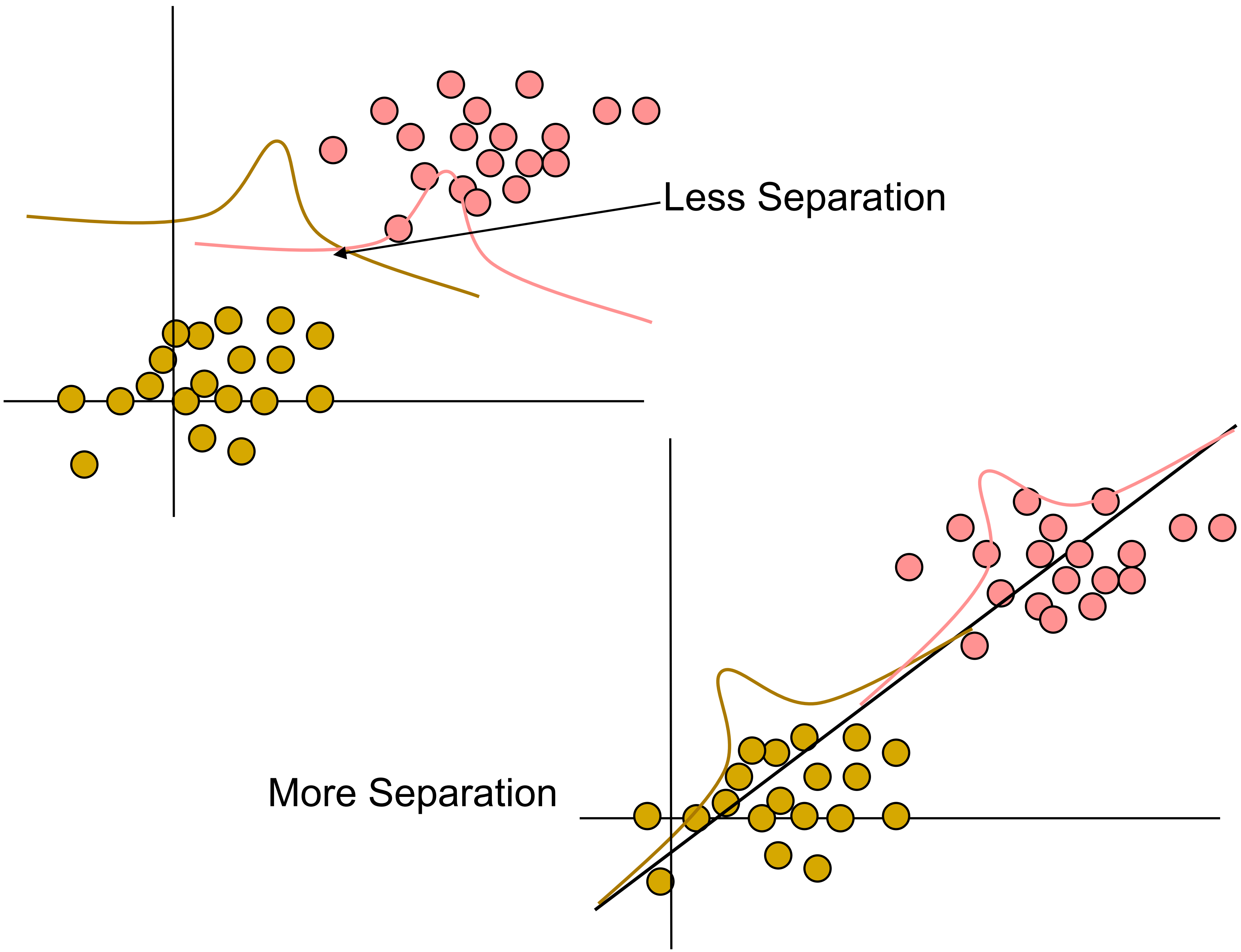
- For a binary classifier

$$\text{Decision Surface } f(x) = 0$$

$$\text{Class } C_0 \text{ if } f(x) > 0$$

$$Class \mathrm{C}_1 \text{ if } f(x) < 0$$

Varying Levels of Separation



Fisher's Linear Discriminant Function

- We want to find the direction (w) of the decision surface such that points are well separated
  - Project points to a line
  - Compute mean and variances for the classes

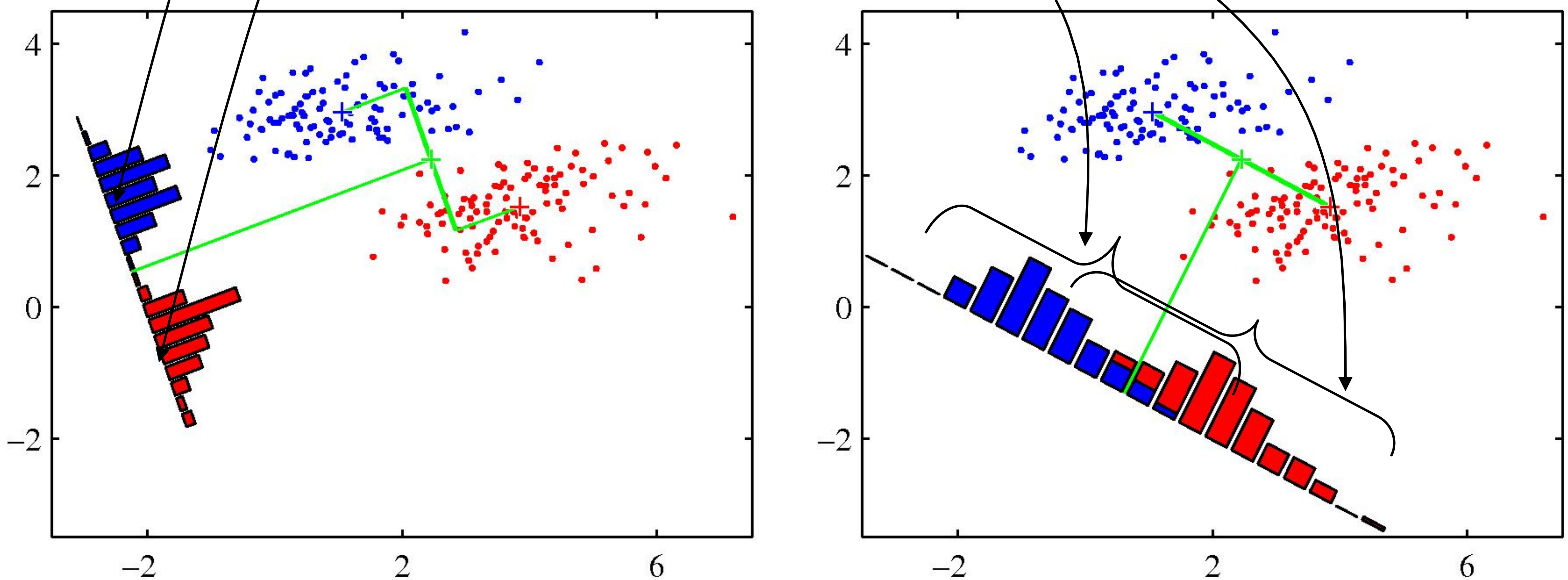
  - Maximize

$$J(w) = \frac{\text{square of separation of projected means}}{\text{Sum of within-class variance}}$$

Maximize a function that will produce large separation between class means (projected) and has smaller within-class variance

Why This Criteria Makes Sense?

$$J(w) = \frac{\text{square of separation of projected means}}{\text{Sum of within-class variance}}$$



Fisher's Linear Discriminant

Projected means

$$J(\mathbf{w}) = \frac{(\mathbf{w}^T \mu_1 - \mathbf{w}^T \mu_0)^2}{\mathbf{w}^T (\sum_1 + \sum_0) \mathbf{w}}$$

$$\mathbf{w} = (\sum_1 + \sum_0)^{-1} (\mu_1 - \mu_0)$$