STEVENS
INSTITUTE of TECHNOLOGY
School of Business

Amir H Gandomi; PhD
Assistant Professor
Stevens Institute of Technology
a.h.gandomi@stevens.edu

Multivariate Data Analysis – BIA 652

Class 4 – Linear Discriminant Analysis & K Nearest Neighbor

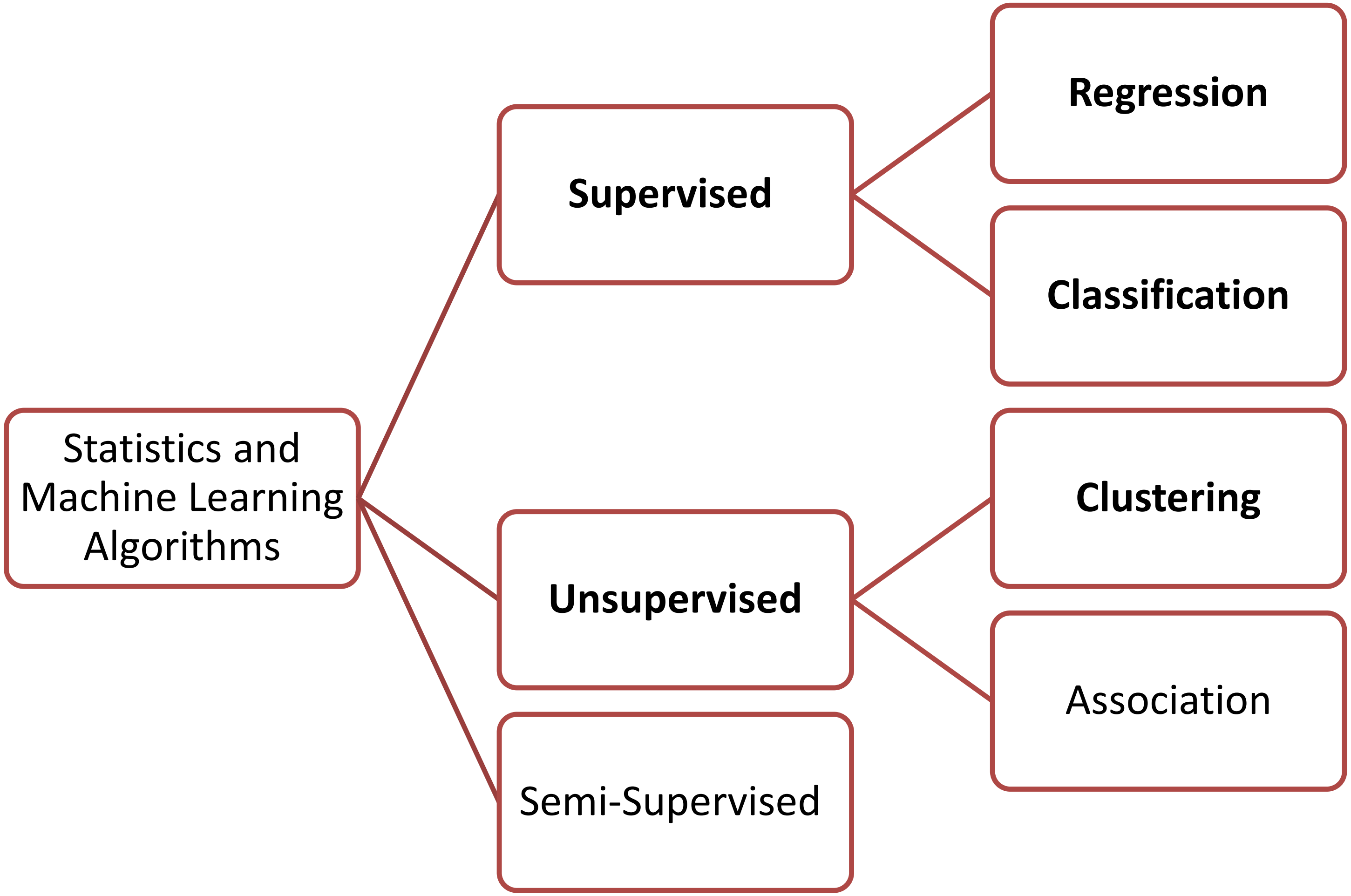
Project Proposal

- Submit the tentative Abstract of Project Proposal next class:
  - Title
  - Names
  - Details about the database
    - Problem
    - Type: Classification (preferred), Regression, Clustering, etc.
    - Inputs and an Output
    - Reference: own (preferred) or borrowed

Overview of Class 5

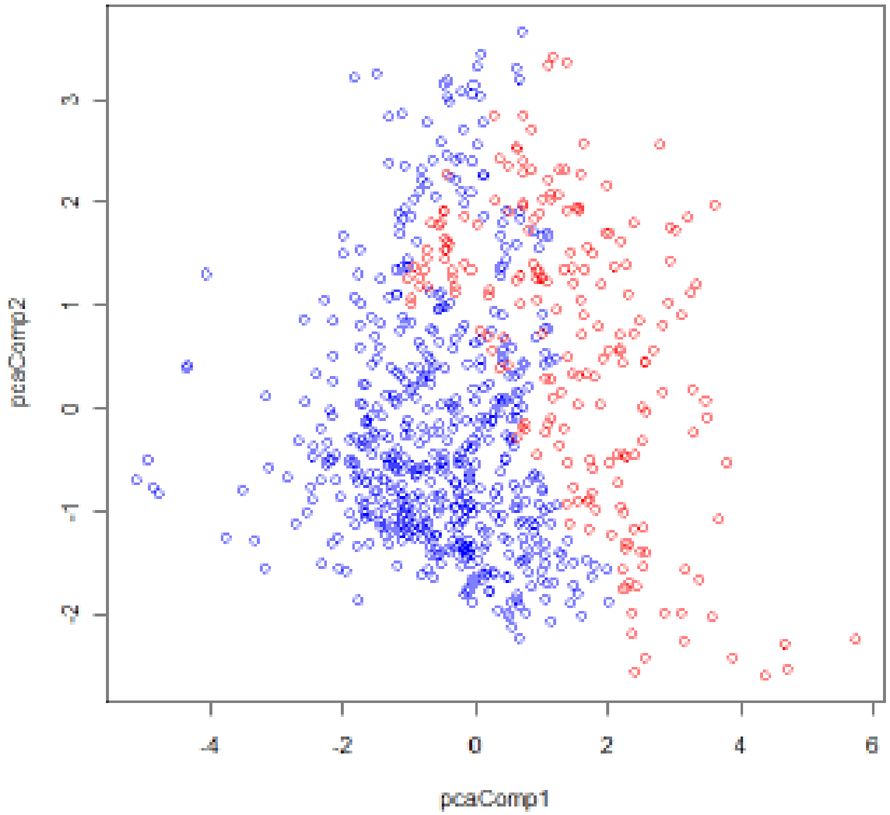
- Start Classification – Chapter 11 & 12
  - Linear Discriminant Analysis
  - K Nearest Neighbor
- Review Problem 11.1 + Discriminant Analysis Chemical
- Homework for Next Class will be posted on Tuesday:
  - A classification problem
- Next class
  - Logistic Regression 1 – Chapter 12
  - Naïve Bayes Classification

Supervised and Unsupervised Algorithms



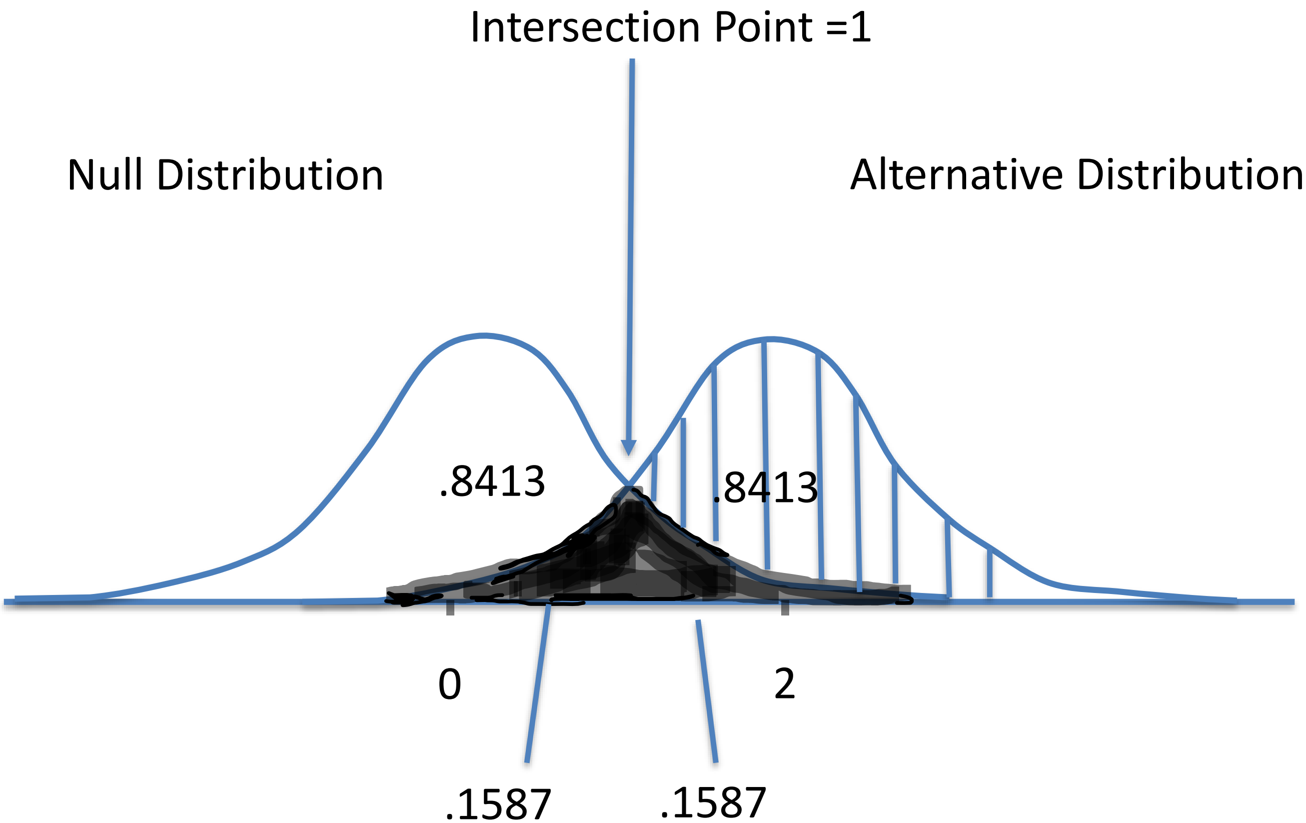
Classification

The Challenge is to be able to “classify” a new member, hence to separate the classes



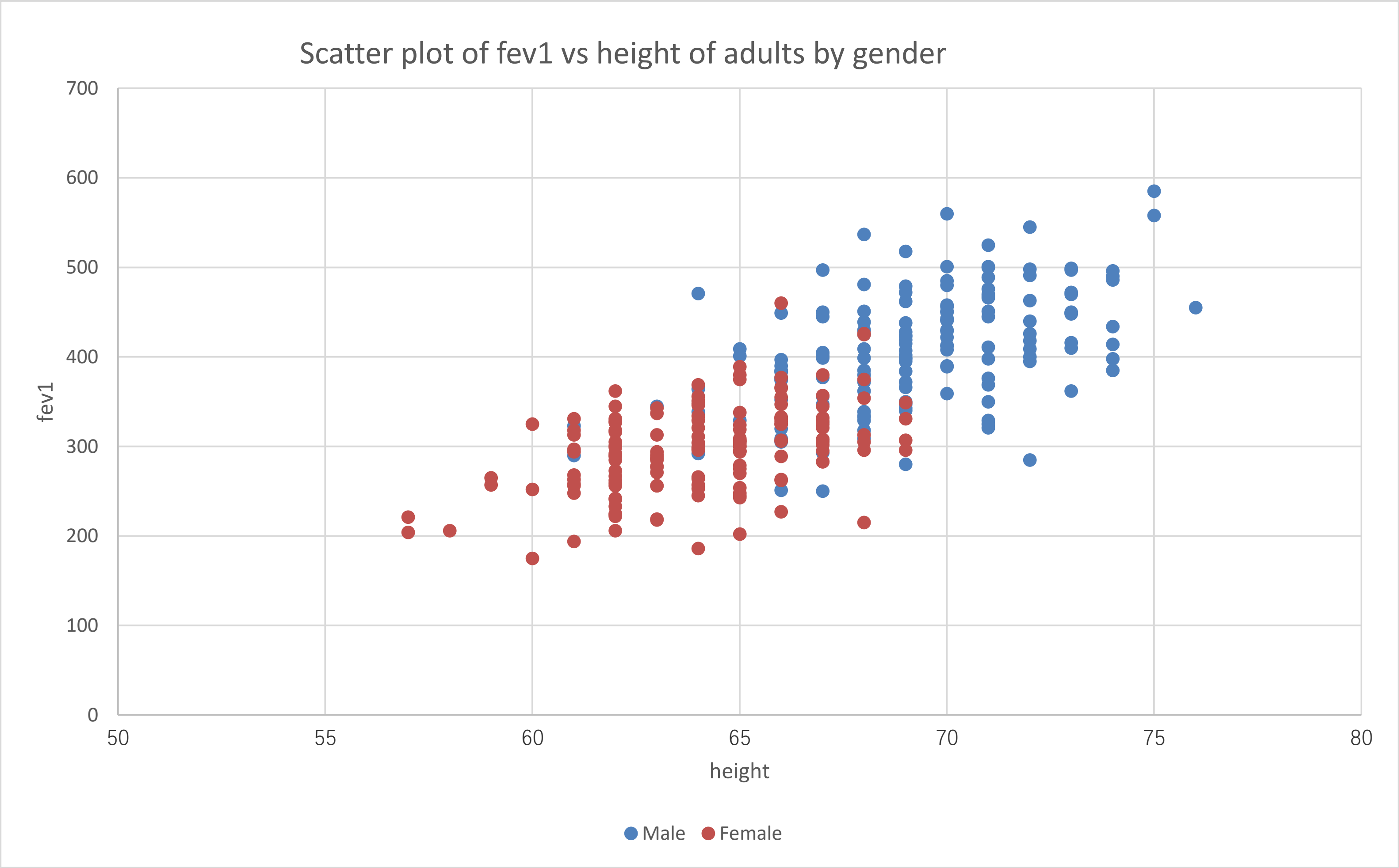
Blue Dots are Non-Diabetic, Red are Diabetic

Classification

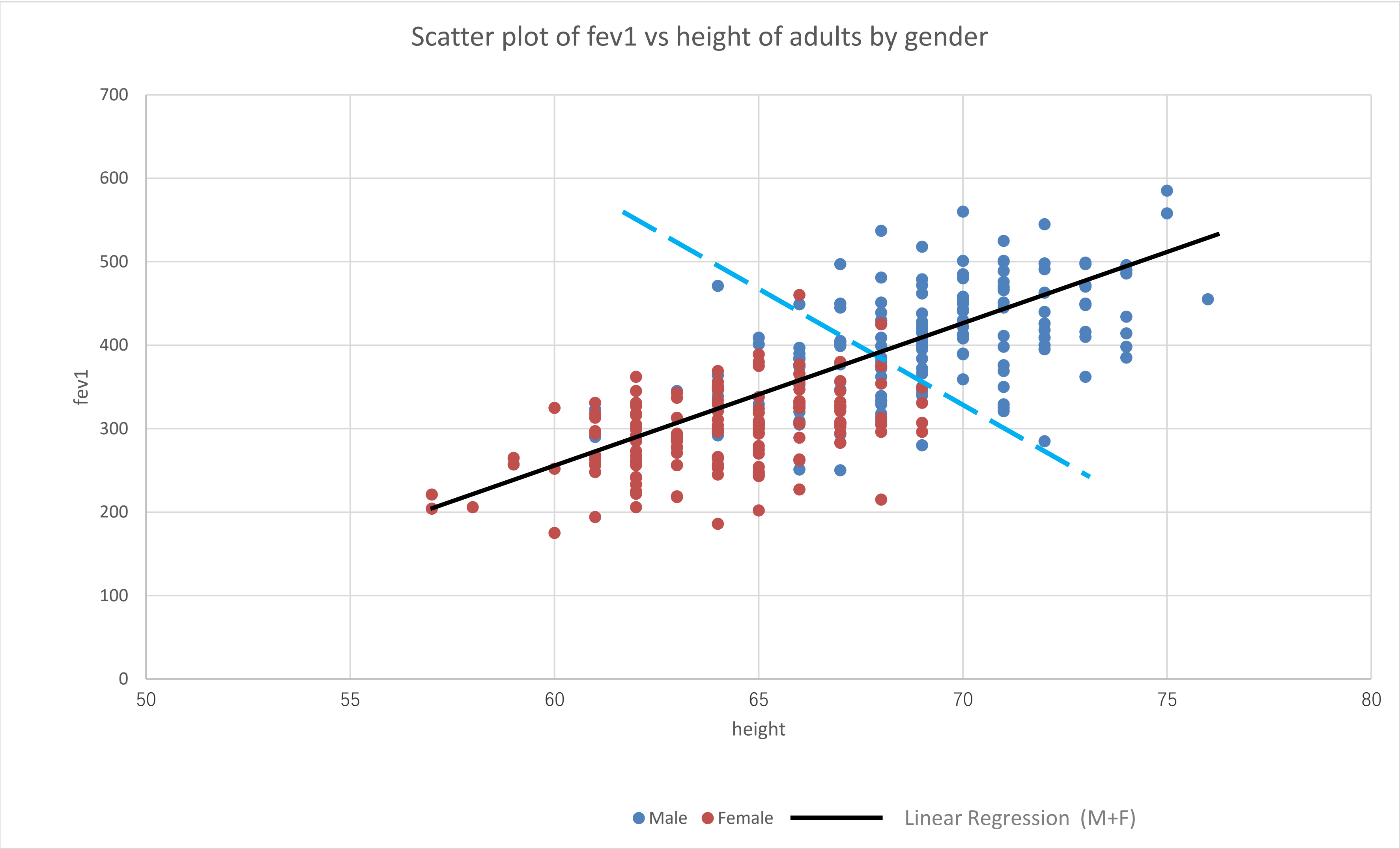


Source of Picture:
http://psychology.okstate.edu/faculty/jgrice/personalitylab/methods.htm

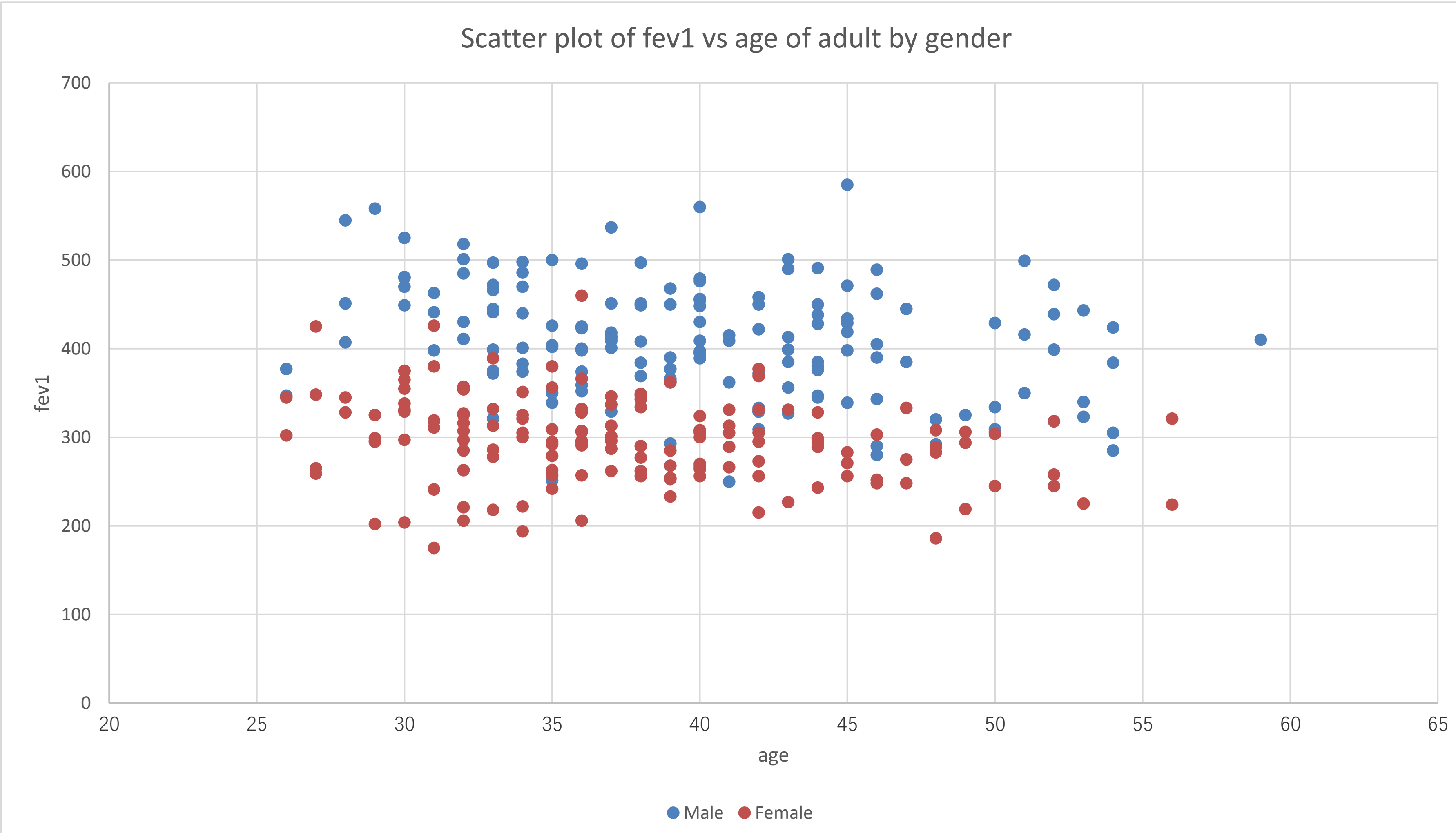
FEV1 vs height



FEV1 vs height – Regression lines

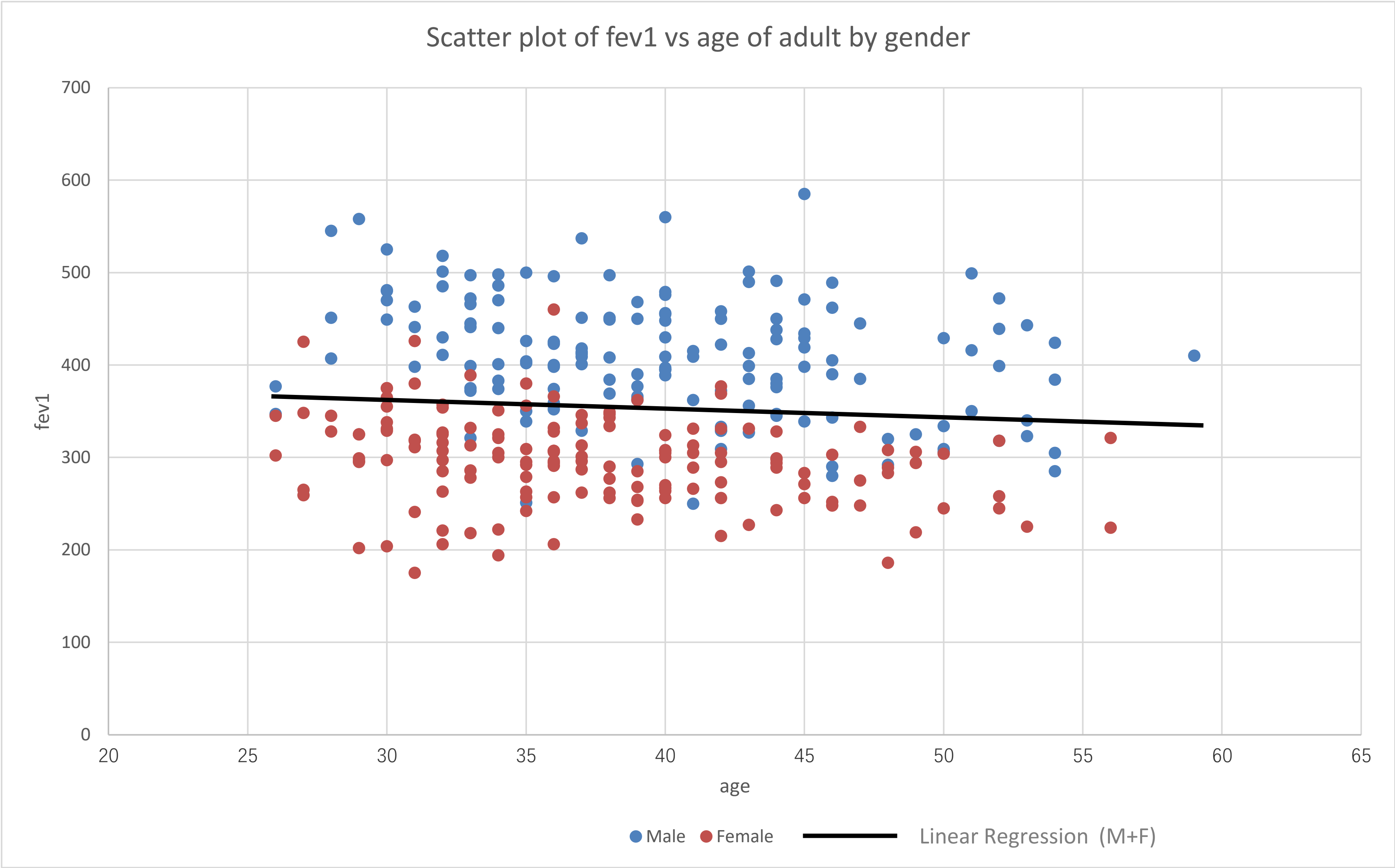


FEV1 vs age



Scatter plot of fev1 vs age of adult by gender
fev1
700
600
500
400
300
200
100
0
20
25
30
35
40
45
50
55
60
65
age
Male
Female

FEV1 vs age– Regression lines



Classification of Classifiers

Classifiers

Discriminative	**Generative**
Learn differences	Learn each class
Model P(y|x)	Model P(x,y)
Such as: LR, KNN	Such as: NB

Discriminative classifiers usually perform better when we have enough data

Discriminant Analysis

Discriminant Analysis: Background

A classical statistical technique

Used for classification long before data mining
- Classifying organisms into species
- Classifying skulls
- Fingerprint analysis

And also used for business data mining (loans, customer types, etc.)

Can also be used to highlight aspects that distinguish classes (**profiling**)

Aims of Discriminant Analysis

- Construct an algorithm to classify individuals into one of two or more groups.
- Compute the probability of an individual belonging to a given group.

Bayes Rule
Notation & Theorem

q_1, q_2 = Prior probability of population 1, 2

$f(X|1 \text{ or } X|2)$ = density function in population 1 or 2

$P(1|X)$ = Posterior probability of population 1

= probability that subject belongs to population 1 given X

$P(2|X) = 1 - P(1|X)$

$$P(1|X) = q_1 f(X|1) / (q_1 f(X|1) + q_2 f(X|2))$$
$$= P(X|1)P(1)/P(X)$$

Discriminant Analysis: Conditional Probability

P(Buy & Male)=?
P(Buy | Male)=?
P(Buy)=?
P(Male)=?
P(Female)=?
P(Male | Buy)=?

	Buy	No Buy	
Male	**10%**	**30%**	**40%**
Female	**15%**	**45%**	**60%**
	25%	**75%**	**100%**

P(Buy|Male)=

P((Male|Buy)P(Buy)/(P(Male|Buy)P(Buy)+P(Male|Not Buy)P(Not Buy)) =
(2/5)(1/4)/((2/5)(1/4) + (2/5)(3/4)) =
.1/(.1 + .3) = .25

Discriminant Analysis: Example:

$\pi_1 =$ Buyers

50%

Gender		Young	Middle	Old	
		Age Category			
	Male	15%	15%	10%	40%
	Female	20%	20%	20%	60%
		35%	35%	30%	100%

$\pi_2 =$ not Buyers

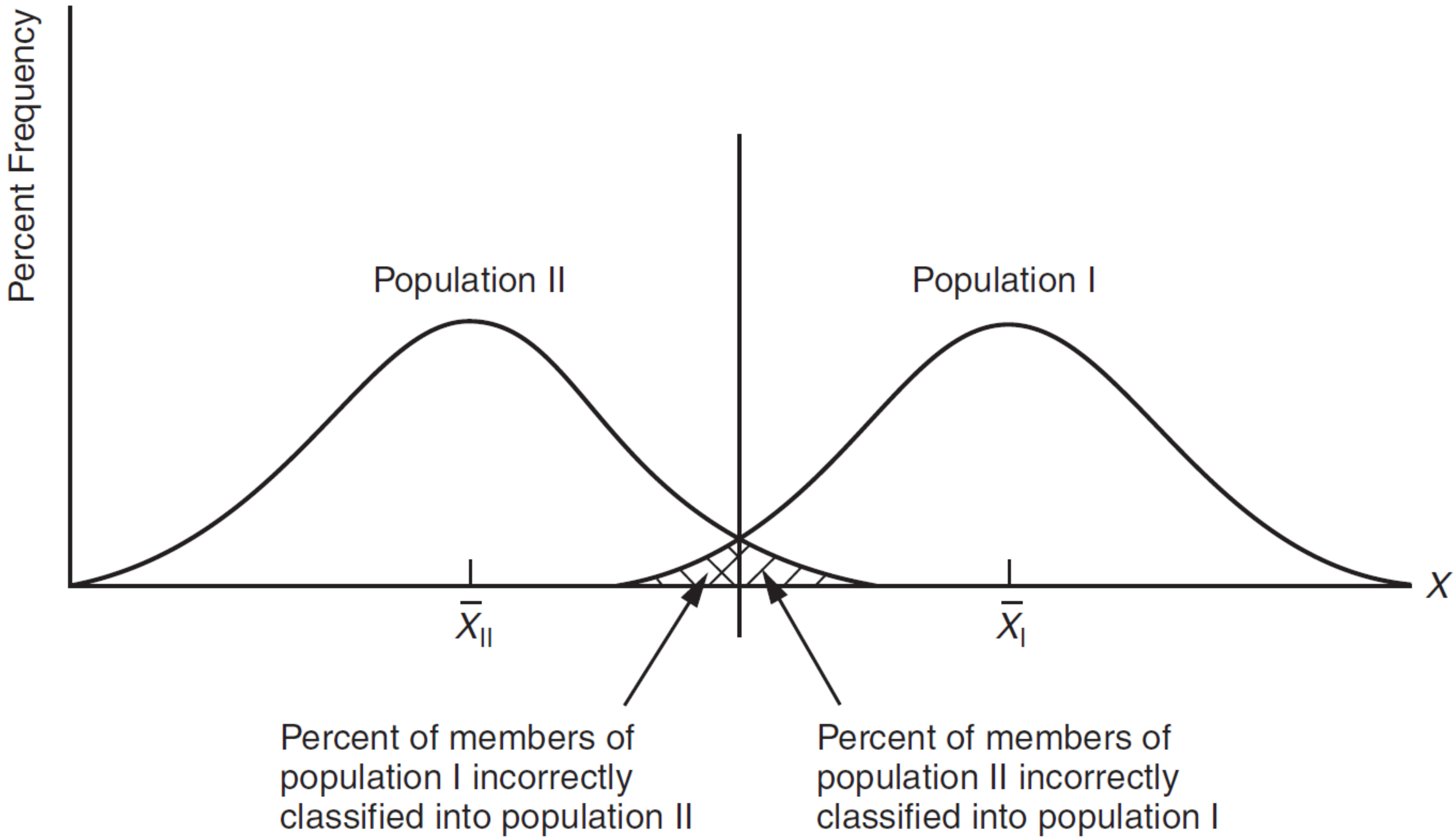
50%

Gender		Young	Middle	Old	
		Age Category			
	Male	20%	20%	20%	60%
	Female	10%	15%	15%	40%
		30%	35%	35%	100%

Discriminant Function

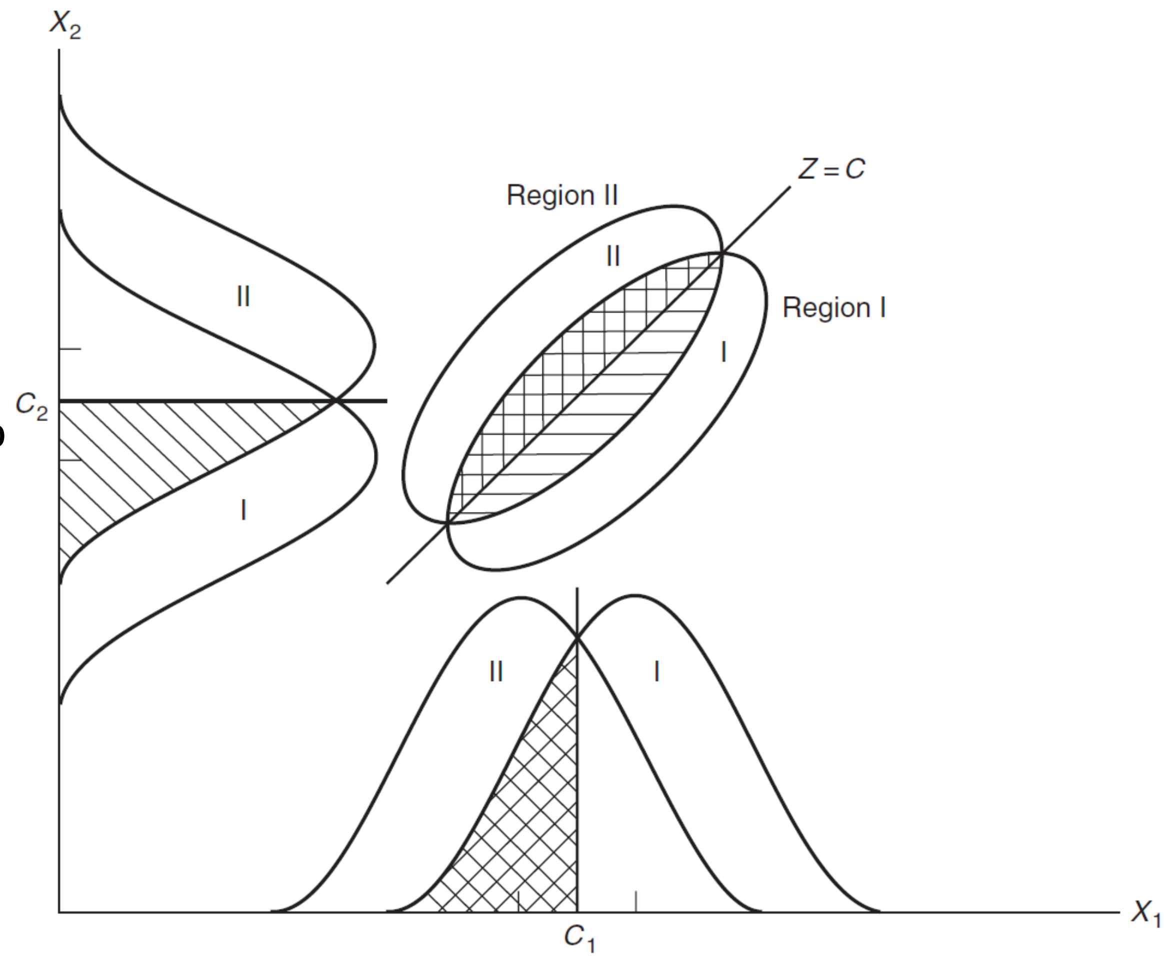
- How do you incorporate different prior probabilities?
- How do you incorporate different costs?
- Based on a training group of individuals whose group membership is known
- Minimizes total probably of misclassification
- Incorporates known prior probabilities, and differing costs of misclassification

Aims of Discriminant Analysis Single Variable, Two Groups (p 243)



Standard Distance & Linear Discriminant Function

Figure 11.3
Classification into Two Groups on the Basis of Two Variables



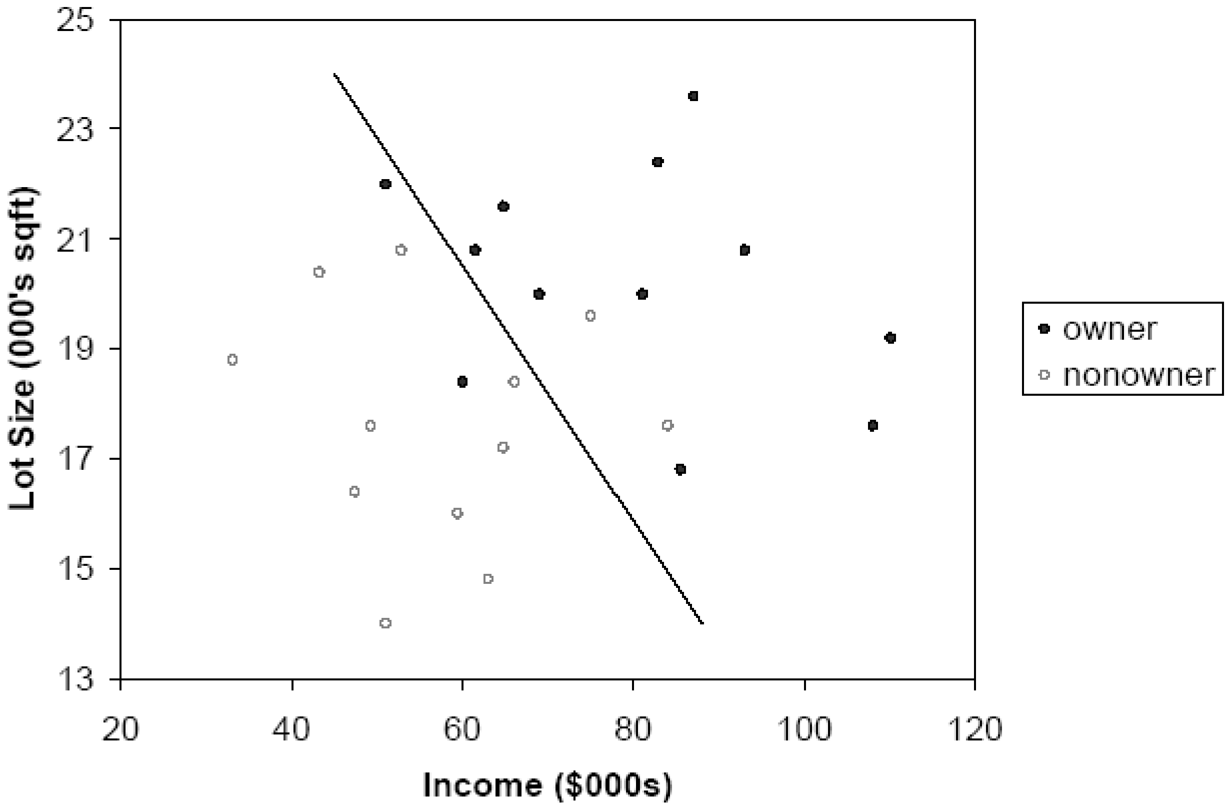
Small Example: Riding Mowers

Goal: classify purchase behavior (buy/no-buy) of riding mowers based on income and lot size

Outcome: owner or non-owner (0/1)

Predictors: lot size, income

Can we manually draw a line that separates owners from non-owners?



Linear Discriminant Function for Two Groups – Part 1 (p 247)

$\mathbf{Z = a_1X_1 + a_2X_2}$

$\bar{Z}_{I}$: Average of Z for Population I
$\bar{Z}_{II}$: Average of Z for Population II

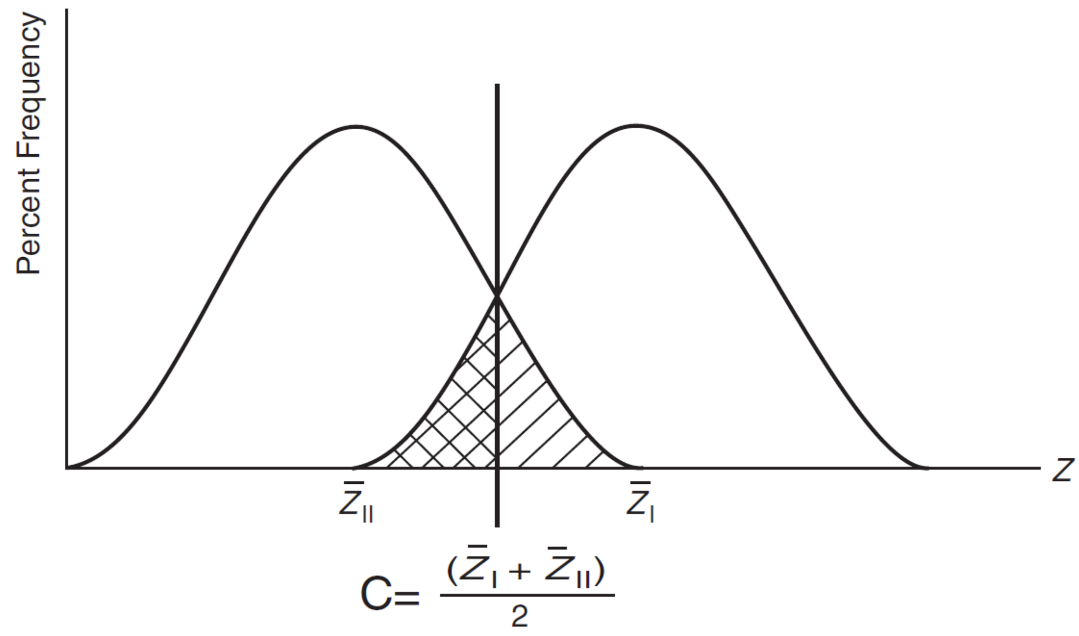


Figure 11.4
Frequency Distributions of *Z* for Populations I and II

Assumptions & Caveats of Discriminant Analysis

1. Assumes multivariate normality of predictors

   When this condition is met, DA is more efficient than other methods (i.e. needs less data to obtain similar accuracy)

   Even when it is not met, DA is robust when we have enough cases in smallest class (> 20). This means it *can be used with dummy variables*!

2. Assumes correlation among predictors within a class is the same across all classes. (Compare correlation tables of each class by eye.)
3. Sensitive to outliers

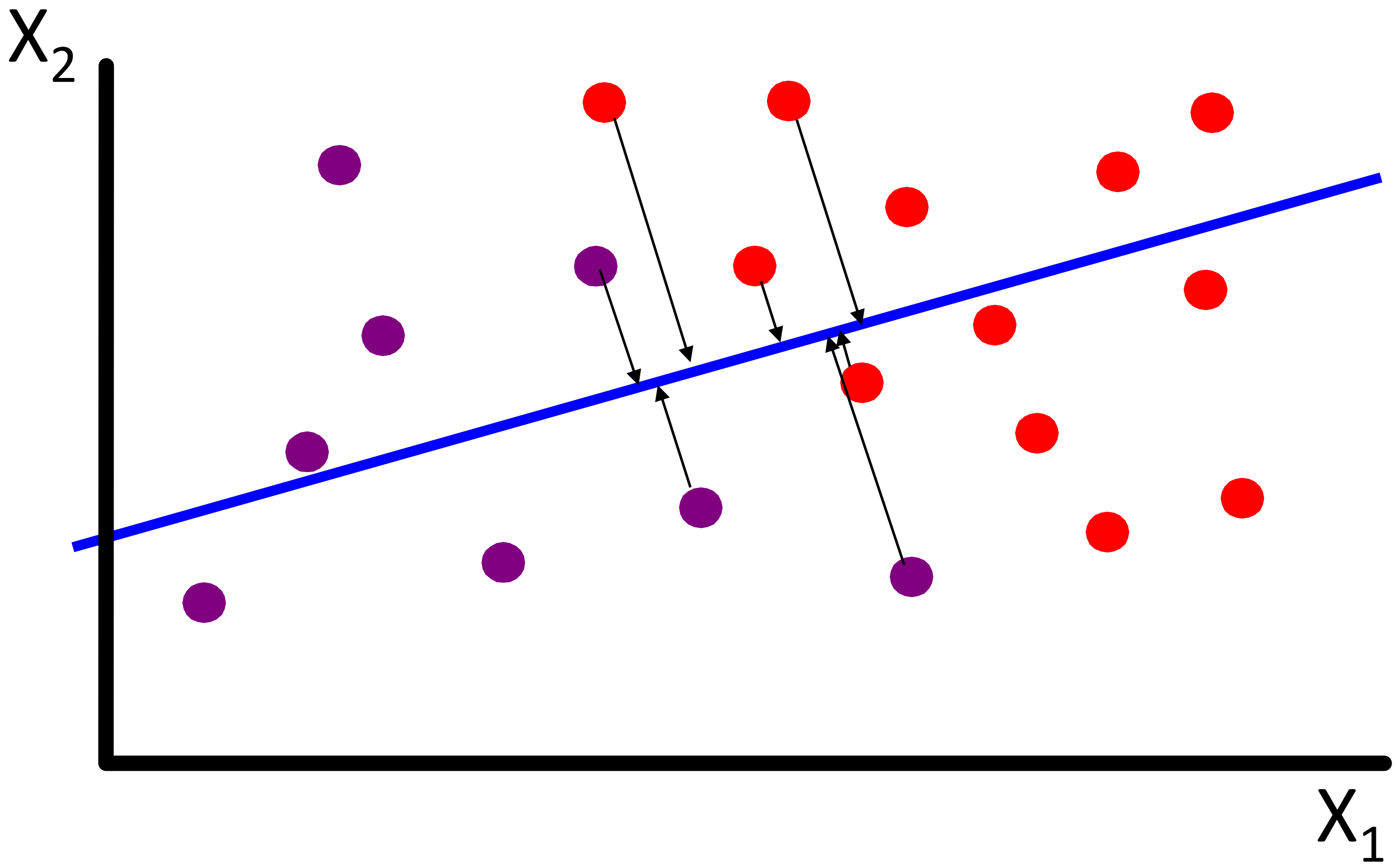
Algorithm for Discriminant Analysis

The Idea

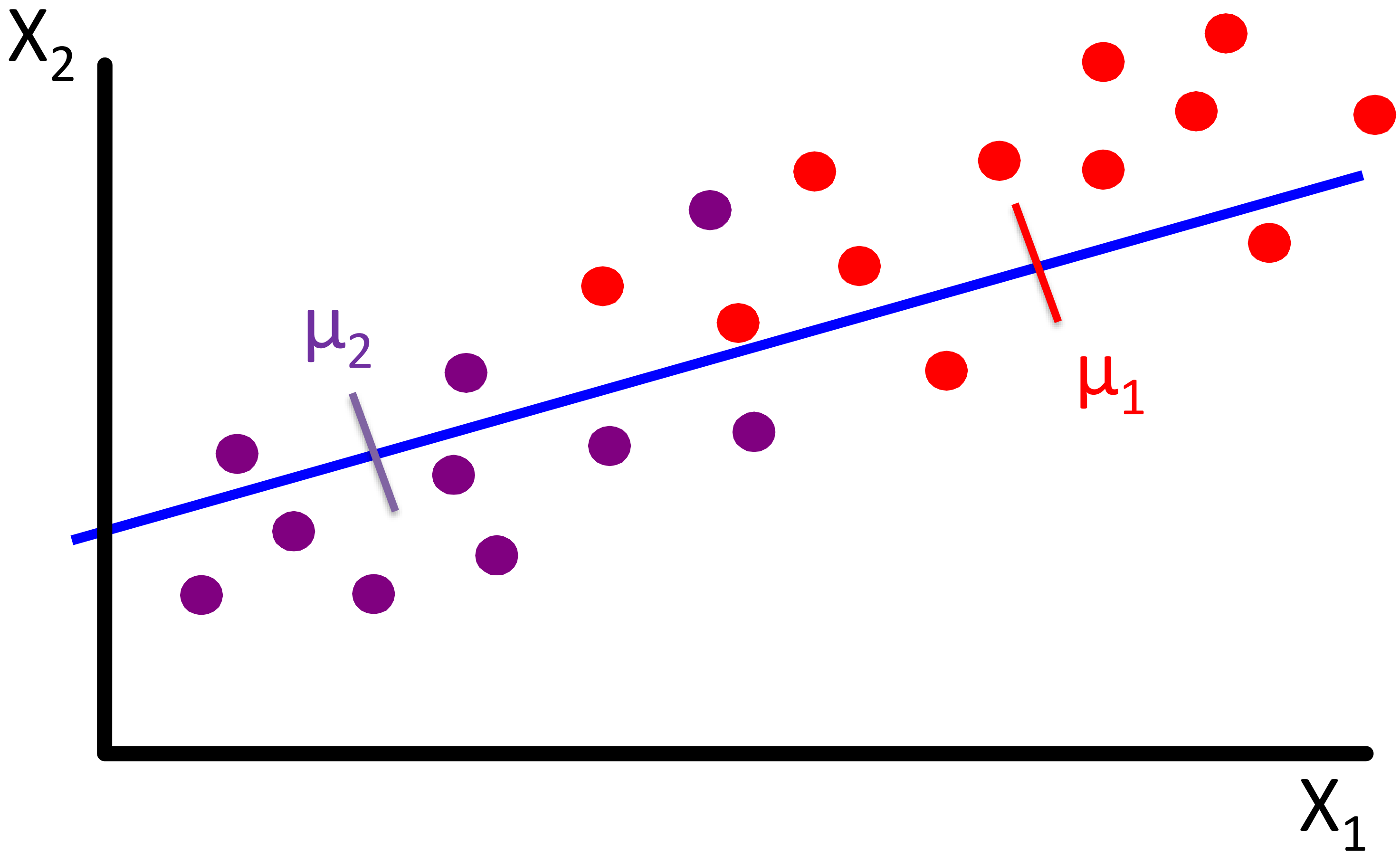
To classify a new record, measure its distance from the center of each class

Then, classify the record to the closest class

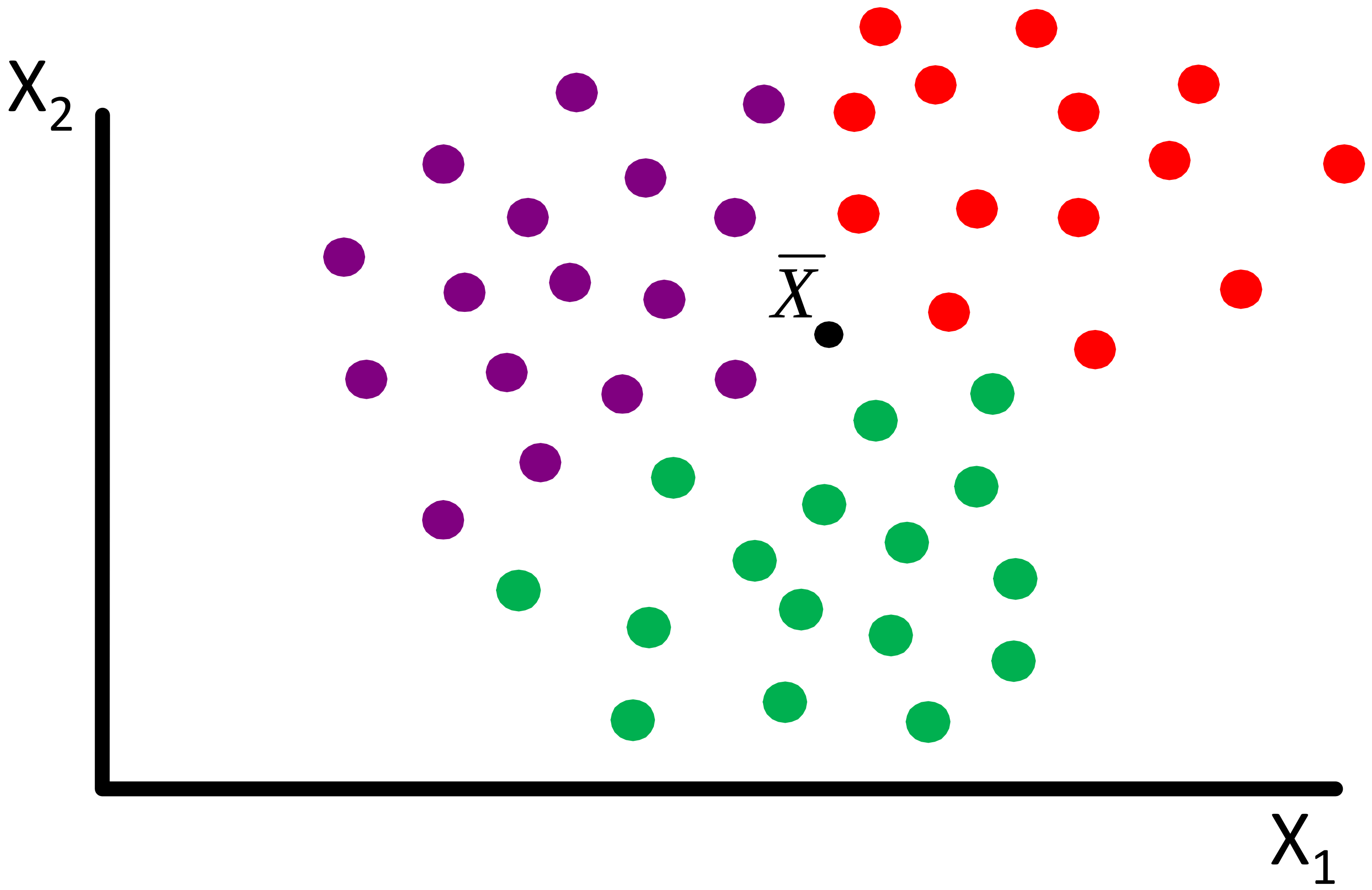
Reducing the dimension



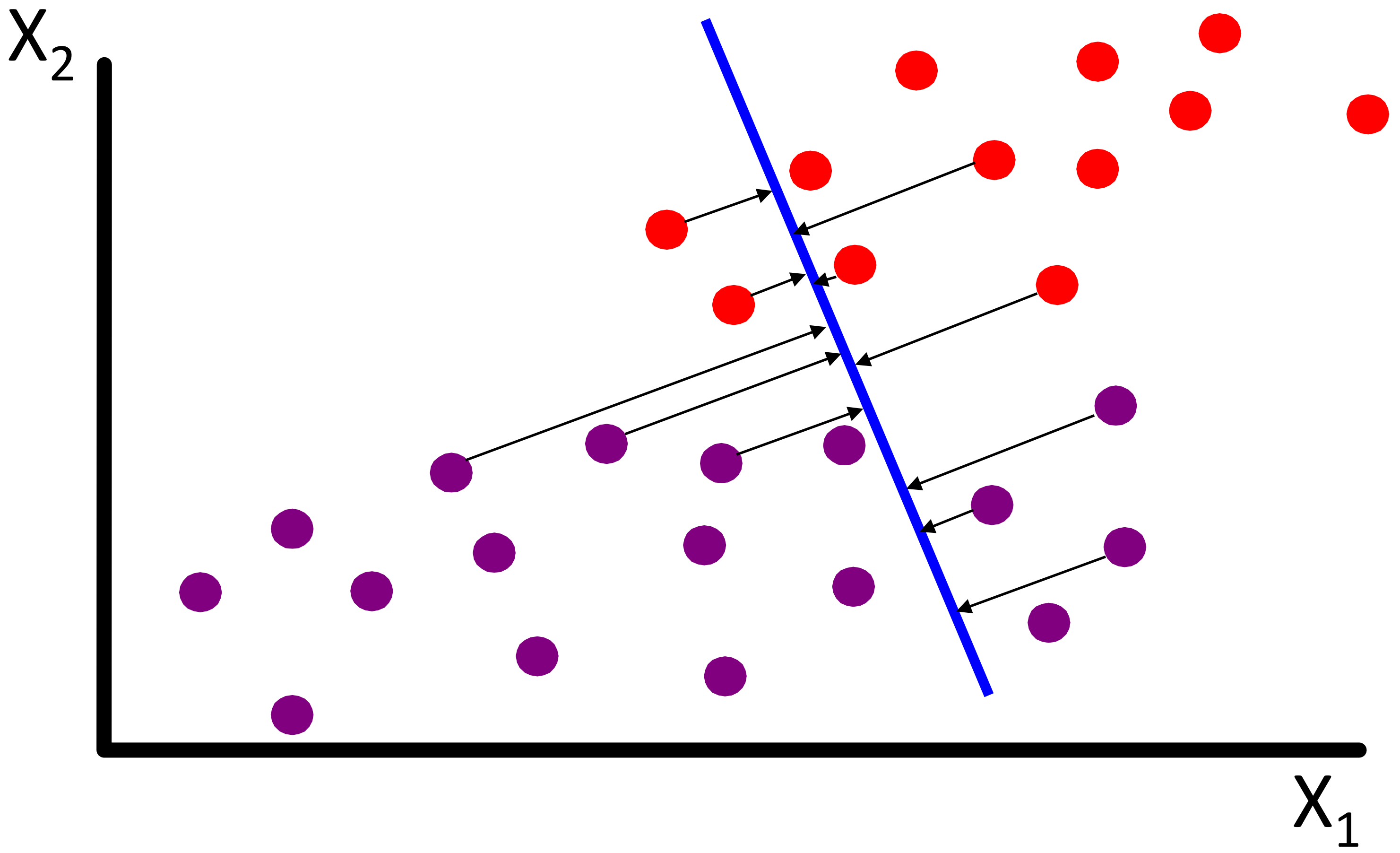
Objective 1: Maximizing Distance between Means (binary class)



Objective 1: Maximizing Distance between Means (>2 classes)



Objective 2: Minimizing Variation



Step 1: Measuring Distance

Need to measure each record's distance from the center of each class

The center of a class is called a *centroid* $\overline{X}$

The centroid is simply a vector (list) of the means of each of the predictors. This mean is computed from all the records that belong to that class.

Step 1: Measuring Distance – cont.

A popular distance metric is Euclidean Distance. We can use it to measure the distance of a record from a class centroid:

$$D_{Eucidean}(X,\bar{X}) = \sqrt{(x_1 - \bar{x}_1)^2 + \cdots + (x_p - \bar{x}_1)^2}$$

Drawbacks:

- Sensitive to scale, variance (can be normalized to be correct)
- Ignores correlation between variables

Instead, use "Statistical (Mahalanobis) Distance"

transpose (convert column to row)

$$D_{Statistical}(X, \bar{X}) = [X - \bar{X}]' S^{-1} [X - \bar{X}]$$

inverse of covariance matrix S

(p-dimension extension of division)

- For a single predictor (p=1), this reduces to a **z-score**
- When p > 1, statistical distance takes account of correlations among predictors (z-score doesn't)

Step 2: Classification Functions

The idea is to create classification score that reflects the distance from each class

This is done by estimating "*classification functions*", which are a function of the statistical distances.

The estimation maximizes the ratio of between-class to within-class variability

Fisher's linear classification functions: one for each class. Used to compute a classification score.

Classify a record to class with highest score

Step 3: Converting to Probabilities

It is possible to convert classification scores to probabilities of belonging to a class:

$$P = \frac{e^{c_k(i)}}{e^{c_i(i)} + e^{c_2(i)} + \cdots + e^{c_m(i)}}$$

Probability that record *i* (with predictor values x_1, x_2, …x_m) belongs to class *k*

The probability is then compared to the cutoff value in order to classify a record

Linear Discriminant Function for Two Groups – Part 1

- Let Z be a linear combination of X (a vector of p variables)

  $Z = a_1 X_1 + a_2 X_2 + \ldots + a_p X_p$

Choosing a_i 's:

- Compute group means and pooled SD of Z based on a_i 's
- Select coefficients to maximize:

  $(\text{mean}Z_1 - \text{mean}Z_2)^2 / s_Z^2$

Linear Discriminant Function for Two Groups – Part 1
Choosing a_i 's (cont)

The vector $\mathbf{a} = (\, a_1 \quad \ldots \quad a_p \,)^T$ **is:**

$$\mathbf{a} = S^{-1} \left(\text{mean}(X_1) - \text{mean}(X_2) \right)$$

Where:

S = pooled covariance matrix, and

mean(X_i) = sample mean for group *I*

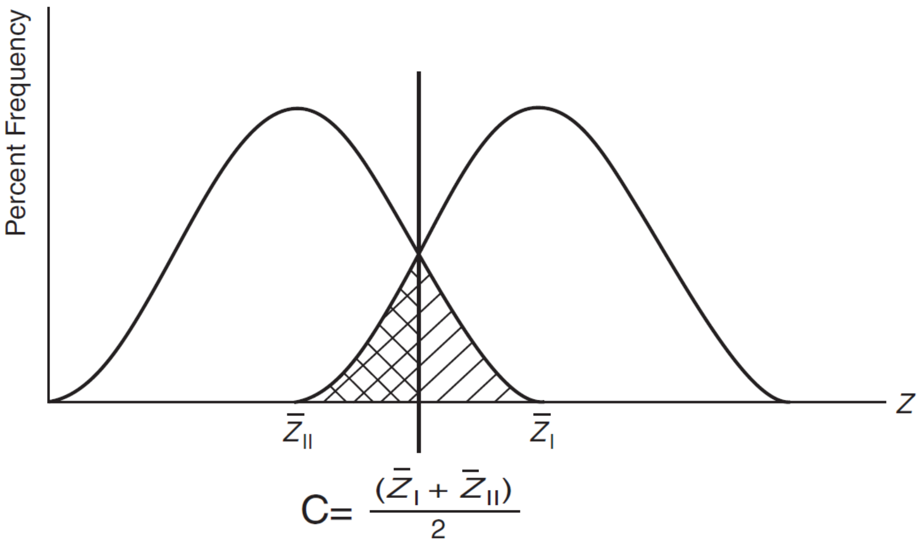
Pooled covariance matrix is estimating variance of several populations with different population mean

Linear Discriminant Function for Two Groups – Part 1
Classification Rule

Place the individual j in:

- Group 1 if $Z \geq C$
- Group 2 if $Z < C$

Where C = the midpoint between the two group means of Z.



Standard Distance & Linear Discriminant Function

- Multivariate Distance:

$$D(m(x_1),m(x_2)) = \{(m(x_1)-m(x_2))^T S^{-1} (m(x_1)-m(x_2))\}^{1/2}$$

$$a = S^{-1} (m(x_1)-m(x_2)) = S^{-1} \times d$$

(or any vector proportional to b)

$$S = \frac{1}{N_1 + N_2 - 2} [(N_1 - 1) \times S_1 + (N_2 - 1) \times S_2]$$

$Z = a_1 \times x_1 + a_2 \times x_2$

C = Mean (Z) = (Mean X_1 + Mean X_2)/2

Assign to: 1 if $Z \geq C$

Assign to: 2 if $Z < C$

Midges Dataset

Species	Antenna Length (mm)	Wing Length (mm)
Af	1.38	1.64
Af	1.4	1.7
Af	1.24	1.72
Af	1.36	1.74
Af	1.38	1.82
Af	1.48	1.82
Af	1.54	1.82
Af	1.38	1.9
Af	1.56	2.08
Apf	1.14	1.78
Apf	1.2	1.86
Apf	1.18	1.96
Apf	1.3	1.96
Apf	1.26	2
Apf	1.28	2

Standard Distance & Linear Discriminant Function (example) - 1

Two Types of Midges (an insect, Af, Apf)
x_1 = Antenna Length $\quad$ x_2 = Wing Length

- $S_1 = \begin{matrix} 98.00 & 80.83 \\ 80.83 & 168.78 \end{matrix}$ & $S_2 = \begin{matrix} 39.47 & 43.47 \\ 43.47 & 77.87 \end{matrix}$ $\rightarrow$ $S = \begin{matrix} 75.49 & 66.46 \\ 66.46 & 133.81 \end{matrix}$
- $S^{-1} = \begin{matrix} 23.54 & -11.67 \\ -11.67 & 13.28 \end{matrix} \times 10^{-3}$
- $d = (m(x_1) - m(x_2)) = \begin{matrix} 18.67 \\ -12.22 \end{matrix}$
- $a = S^{-1} \times d = \begin{matrix} 0.582 \\ -0.381 \end{matrix}$
- $D = (d^T a)^{1/2} = 3.94$

Standard Distance & Linear Discriminant Function (example) - 2

$$Z = 0.582\, x_1 - 0.381\, x_2$$

Sample Mean Af = 13.507

Sample Mean Apf = -2.01

C = V mean = (13.507-2.01)/2 = 5.75

Assign to: Af if $Z \geq 5.75$

Apf if $Z < 5.75$

Standard Distance & Linear Discriminant Function (example in the book, P. 252)

	Not Depressed	Depressed	
Avg age	45.2	18.1	
Avg Income	16.68	15.98	
Age	0.1634	0.1425	a_1 = 0.0209
Income	0.136	0.1024	a_2 = 0.0336
Constant	-5.8641	-4.3483	C = 1.5158

Example: Loan Acceptance

In the prior small example, separation is clear.

In data mining applications, there will be more records, more predictors, and less clear separation.

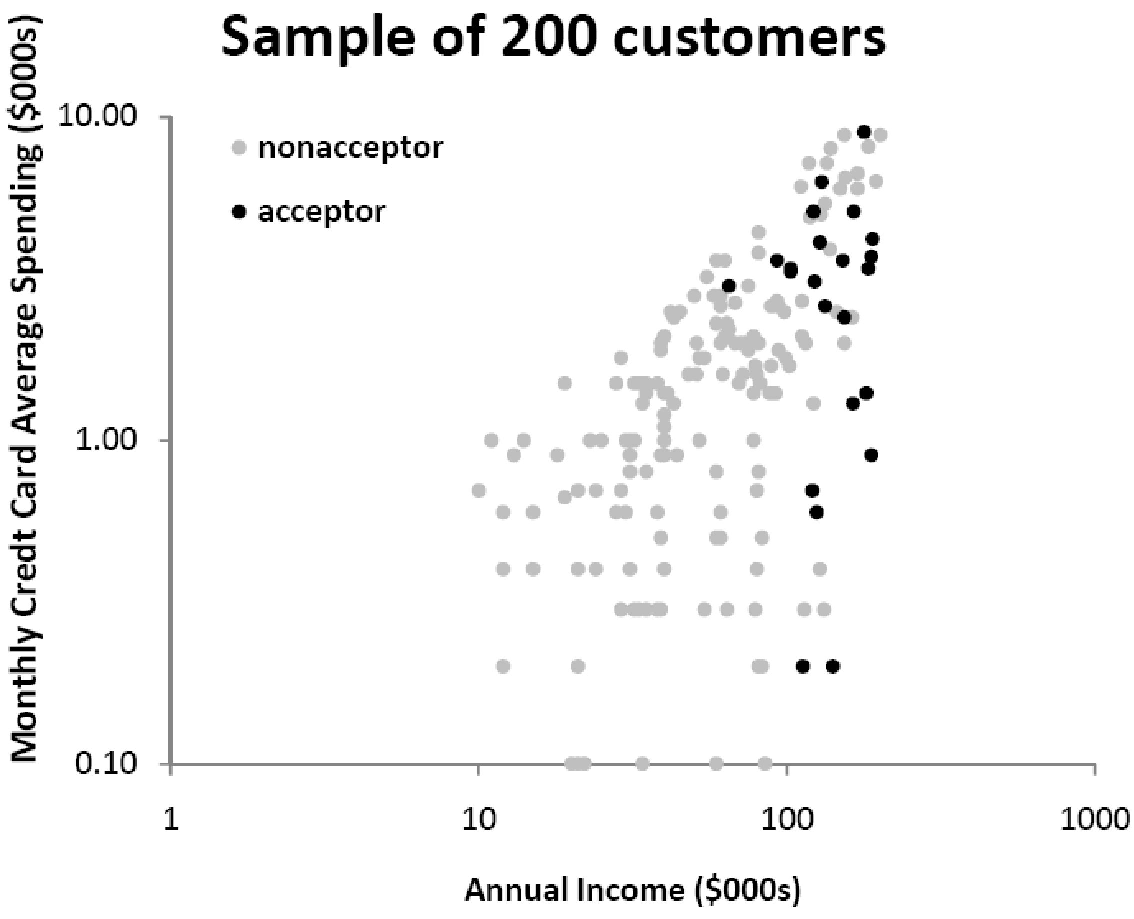
Consider Universal Bank example with only 2 predictors:

Outcome: accept/don't accept loan

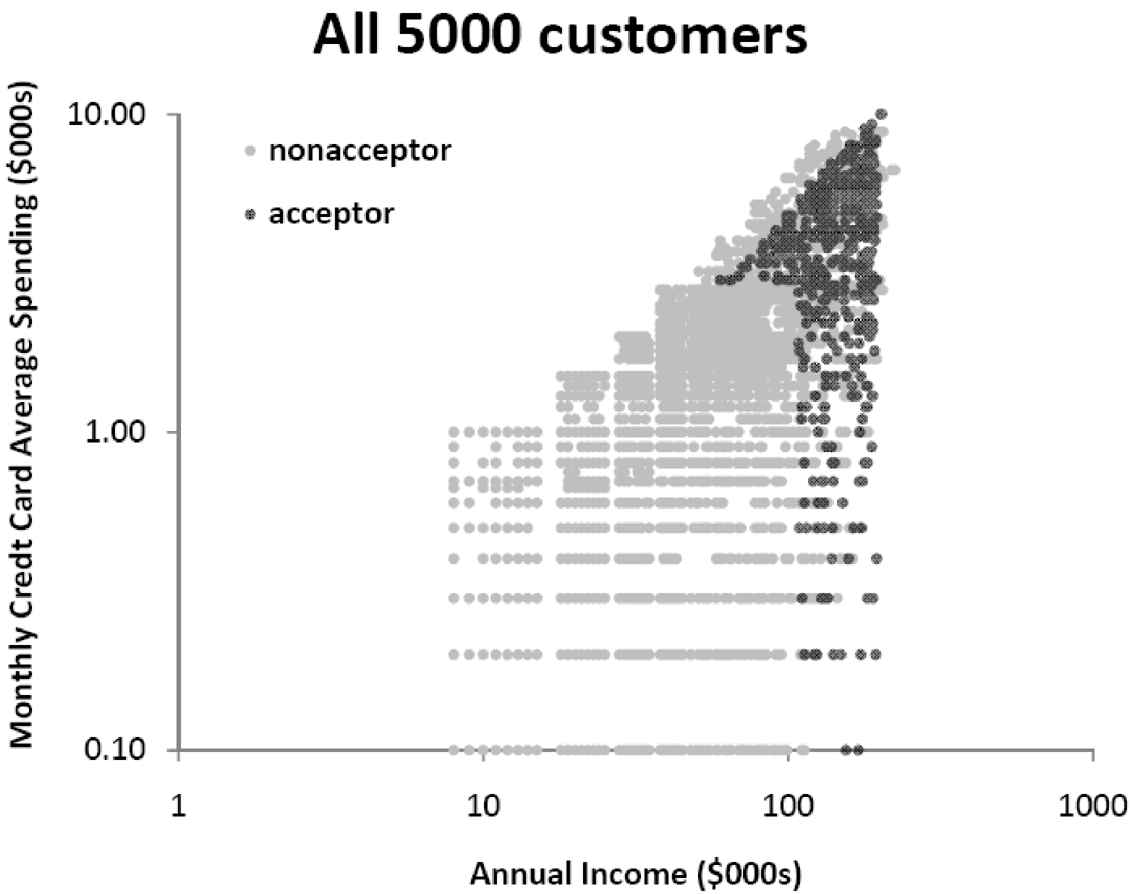
Predictors:

- Annual income (Income)
- Avg. monthly credit card spending (CCAvg)

Sample of 200 customers



5000 customers



Dummy Variables
Nominal Variable > 2 Categories

- Example: Depression Data Set
  - Y = outcome is binary: 0, 1 (e.g. depressed = 1)
  - X nominal: (e.g. religion)
  - Coefficient of D_1 = ln (OR) for "Catholic" vs. "Other"

Religion	D_1	D_2	D_3
Catholic	1	0	0
Protestant	0	1	0
Jewish	0	0	1
Other	0	0	0

Improving Classifications

Prior Probabilities

If classes are not equally frequent, or their frequency in the sample does not reflect reality, then classification functions can be improved

How?

Incorporate prior (or real) probabilities of class membership:

•e.g. Add $log(p_j)$ to the classification function for class j

P_j is probability of a case belongs to class j

Adjusting the Dividing point

- C = $\frac{\bar{z}_I + \bar{z}_{II}}{2}$ + K

- If the number of data are different

$$K = \ln \frac{\bar{q}_{II}}{\bar{q}_I}$$

- Assuming falsely labeling data have different importance:
- e.g.: I is four times as serious to falsely labeled
  - $cost(II\ given\ I)$= 1
  - $cost(I\ given\ II)$= 4

$$K = \ln \frac{\bar{q}_{II}.\ cost(I\ given\ II)}{\bar{q}_I.\ cost(II\ given\ I)}$$

Posterior Probabilities

- Probability of belonging to Pop I

$$= \frac{1}{1+\exp(-z+c)}$$

Linear Regression Analogy

- Assuming DV is dummy variables as:

  Y=$N_{II}/(N_I+N_{II})$ From population 1

  Y=$-N_I/(N_I+N_{II})$ From population 2

- Then regression coefficients are proportional to discriminant function coefficients

- Also:

$$D^2=\frac{R^2}{1-R^2}\,\frac{(N_I+N_{II})\,(N_I+N_{II}-2)}{(N_I N_{II})}$$

Testing Variable Contribution & Variable Selection

- Testing the Null Hypothesis that:
  - Two population means for each variable is identical (population $D^2 = 0$)
  - F statistic value for each pair:

$$F = \frac{N_I + N_{II} - P - 1}{P(N_I + N_{II} - 2)} \times \frac{N_I N_{II}}{N_I + N_{II}} \times D^2$$

Degrees of Freedom: P, $N_I + N_{II} - P - 1$

- Variable selection is similar to regression
  – Using population D^2 instead of multiple R^2

Summary

Discriminant analysis is based on measuring the distance of a record from the class centers

The distance metric used is statistical distance, which takes into account the correlations between predictors

Suitable for small datasets

Assumptions: **equal correlations within each class**, and normality (but fairly robust to violation of normality)

Sensitive to outliers (explore the data!)

Classification functions useful for profiling: can order predictors in terms of separating the classes

Summary of Classification

- **Purpose: Train on a supervised dataset so that new entries can be "classified" into one of two different group. Techniques for more than two groups available.**
- **Techniques:**
  - **Linear Discriminant Analysis – One of the First Techniques**
    - **Assumptions:**
      - **Multivariate Normality of Predictors**
      - **Correlation of Predictors the same across classes**
      - **Sensitive to Outliers**

  - **K Nearest Neighbor – Simple and Often Accurate Classifier**
    - **Non Parametric**
    - **Requires concept of distance**
    - **In SAS, part of proc discrim**

K Nearest Neighbor

K Nearest Neighbor

- A case is classified according to the majority vote of its K nearest neighbors.
- It is then given the class most common among these neighbors
- Obviously we need a concept of “nearness”
- For continuous values we might use:

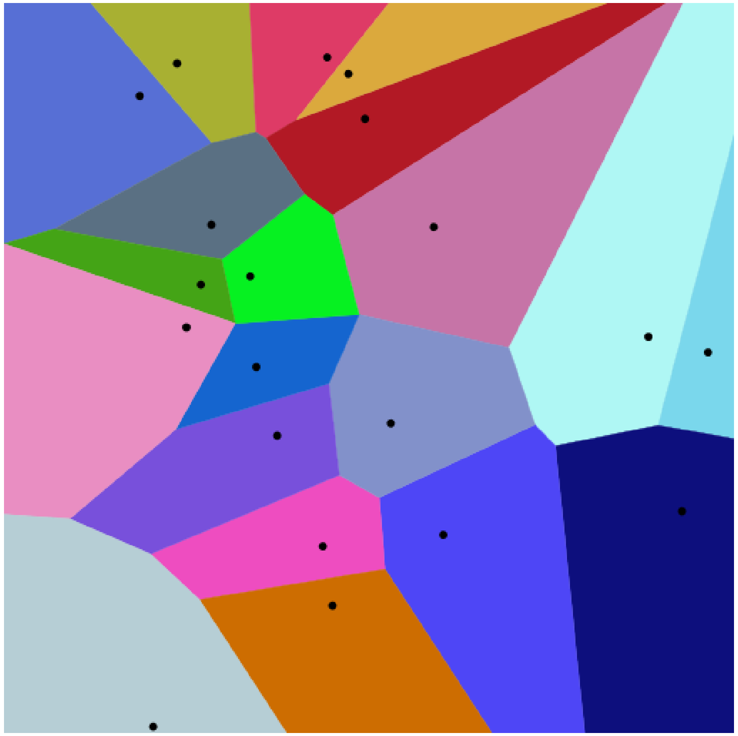
Euclidean Distance $$\sqrt{\sum_{i=1}^{k}(x_i - y_i)^2}$$

Manhattan Distance $$\sum_{i=1}^{k}|x_i - y_i|$$

- For Categorical Variables we might use Hamming Distance:
  - If x = y, D(x-y) = 0
  - If x ≠ y, D(x-y) = 1

Voronoi diagram

– Partitioning of a plane into regions



Euclidean Distance

Manhattan Distance

K Nearest Neighbor

- Choice of K is often done retrospectively or experimentally.
- Generally K between 3 and 10 is used.
- When using distance, Standardization of the variables is often a good idea.
  - Reminder: $x^* = (x-\mu)/\sigma$
- There could be: Ties in voting or in distance (i.e. more than K are the same distance). Dealing with this is often fairly arbitrary.
- Weighting, as in giving more weight to the closest neighbors is another way to deal with these problems.

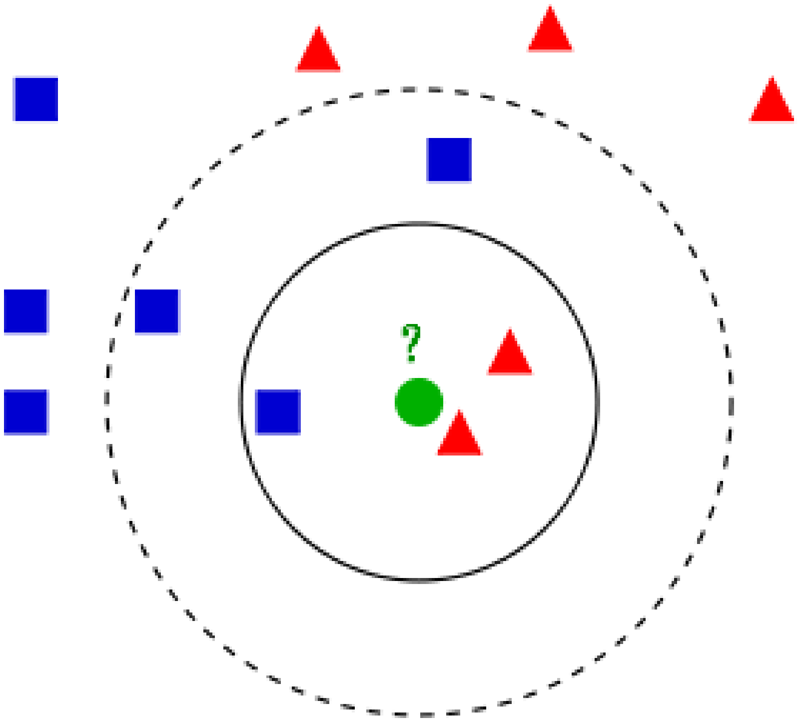
- SAS 12.11 Example

K Nearest Neighbor

- No assumptions about the data
- To classify a given point:
  - Choose a K, often by experimentation
  - Choose a distance metric, e.g. Euclidean n space
  - Find the K nearest predictor points
  - Classify by dominant predictor point values
- Points can be weighted
- Experimentation to choose K is a good idea.

K Nearest Neighbor

- Low K gives strong locality, but high noise
- Higher K provides smoother results, but less locality



K Nearest Neighbor

- 24 Households classified as owning/not owning riding mowers.
- Predictors are Income and Lot Size

Income	Lot_Size	Ownership
60.0	18.4	owner
85.5	16.8	owner
64.8	21.6	owner
61.5	20.8	owner
87.0	23.6	owner
110.1	19.2	owner
108.0	17.6	owner
82.8	22.4	owner
69.0	20.0	owner
93.0	20.8	owner
51.0	22.0	owner
81.0	20.0	owner
75.0	19.6	non-owner
52.8	20.8	non-owner
64.8	17.2	non-owner
43.2	20.4	non-owner
84.0	17.6	non-owner
49.2	17.6	non-owner
59.4	16.0	non-owner
66.0	18.4	non-owner
47.4	16.4	non-owner
33.0	18.8	non-owner
51.0	14.0	non-owner
63.0	14.8	non-owner

K Nearest Neighbor

- Classify By Hand, using K=3, items 2, 14
- See next slide
- Classify new entry (80, 19) by hand.
- See following slide

Income	Lot_Size	Ownership
60.0	18.4	owner
85.5	16.8	owner
64.8	21.6	owner
61.5	20.8	owner
87.0	23.6	owner
110.1	19.2	owner
108.0	17.6	owner
82.8	22.4	owner
69.0	20.0	owner
93.0	20.8	owner
51.0	22.0	owner
81.0	20.0	owner
75.0	19.6	non-owner
52.8	20.8	non-owner
64.8	17.2	non-owner
43.2	20.4	non-owner
84.0	17.6	non-owner
49.2	17.6	non-owner
59.4	16.0	non-owner
66.0	18.4	non-owner
47.4	16.4	non-owner
33.0	18.8	non-owner
51.0	14.0	non-owner
63.0	14.8	non-owner

K Nearest Neighbor

Owner #	Income	Lot Size	Distance to #2	Distance to #14	Neighbor of #2	Neighbor of #14
1	60	18.4	25.550	7.589		
2	**85.5**	**16.8**	**0**	**32.944**		
3	64.8	21.6	21.249	12.027		
4	61.5	20.8	24.331	8.7		
5	87	23.6	6.963	34.314		
6	110.1	19.2	24.716	57.322		
7	108	17.6	22.514	55.293		
8	82.8	22.4	6.217	30.042	Owner	
9	69	20	16.807	16.220		
0	93	20.8	8.5	40.2		
11	51	22	34.890	2.163		Owner
12	81	20	5.522	28.211	Owner	
13	75	19.6	10.867	22.232		
14	**52.8**	**20.8**	**32.944**	**0**		
15	64.8	17.2	20.704	12.528		
16	43.2	20.4	42.453	9.608		
17	84	17.6	1.7	31.364	Non-owner	
18	49.2	17.6	36.309	4.817		Non-owner
19	59.4	16	26.112	8.161		
20	66	18.4	19.565	13.416		
21	47.4	16.4	38.102	6.966		
22	33	18.8	52.538	19.901		
23	51	14	34.613	7.034		Non-owner

K Nearest Neighbor

Classify new entry
(80, 19) :

Owner #	Income	Lot Size	Distance to (80,19)	Neighbor of (18,19)
1	60	18.4	20.009	
2	85.5	16.8	5.924	
3	64.8	21.6	15.421	
4	61.5	20.8	18.587	
5	87	23.6	8.376	
6	110.1	19.2	30.101	
7	108	17.6	28.035	
8	**82.8**	**22.4**	**4.404**	**Owner**
9	69	20	11.045	
0	93	20.8	13.124	
11	51	22	29.155	
12	**81**	**20**	**1.414**	**Owner**
13	75	19.6	5.0359	
14	52.8	20.8	27.259	
15	64.8	17.2	15.306	
16	43.2	20.4	36.826	
17	**84**	**17.6**	**4.238**	**Non-owner**
18	49.2	17.6	30.832	
19	59.4	16	20.817	
20	66	18.4	14.013	
21	47.4	16.4	32.703	
22	33	18.8	47	
23	51	14	29.428	

Summary of Classification (Continued)

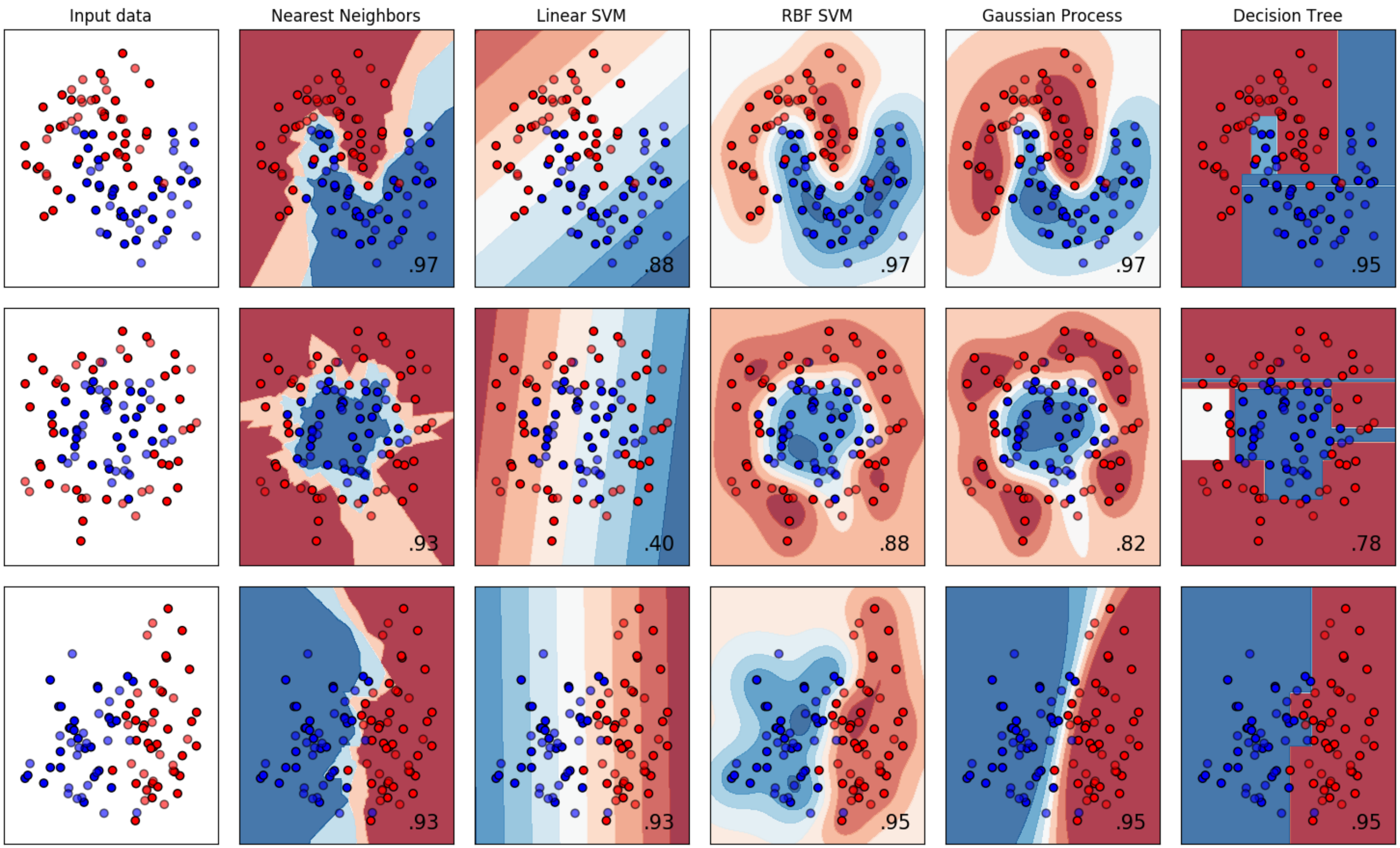
- **Purpose: Train on a supervised dataset so that new entries can be "classified" into one of two different group. Techniques for more than two groups available.**
- **Techniques:**
  - **Logistic Regression – One of Most Widely Used Techniques Across Many Fields. In some fields the Standard.**
    - **Based on a linear function of Odds.**
    - **Directly estimates the parameters of P(Y|X).**
    - **Typically a good option if there is a lot of noise, moderate dimensionality, and a reasonable amount of data.**
    - **Provides a Probabilistic Classifier.**
  - **(Gaussian) Naïve Bayesian – Simple and Often Very Accurate Classifier**
    - **Uses all variables, but treats them individually**
    - **Assumes that Independent Variables are Conditionally Independent given Y.**
    - **Based on Bayes Rule of Conditional Probability**
    - **Directly estimates the parameters of P(Y) and P(X|Y)**
    - **GNB and Logistic Regression Asymptotically Converge to Same Classifier.**
  - **Other Options: Support Vector Machines (SVM), Decision Tree, etc.**

A simple program for LDA and KNN

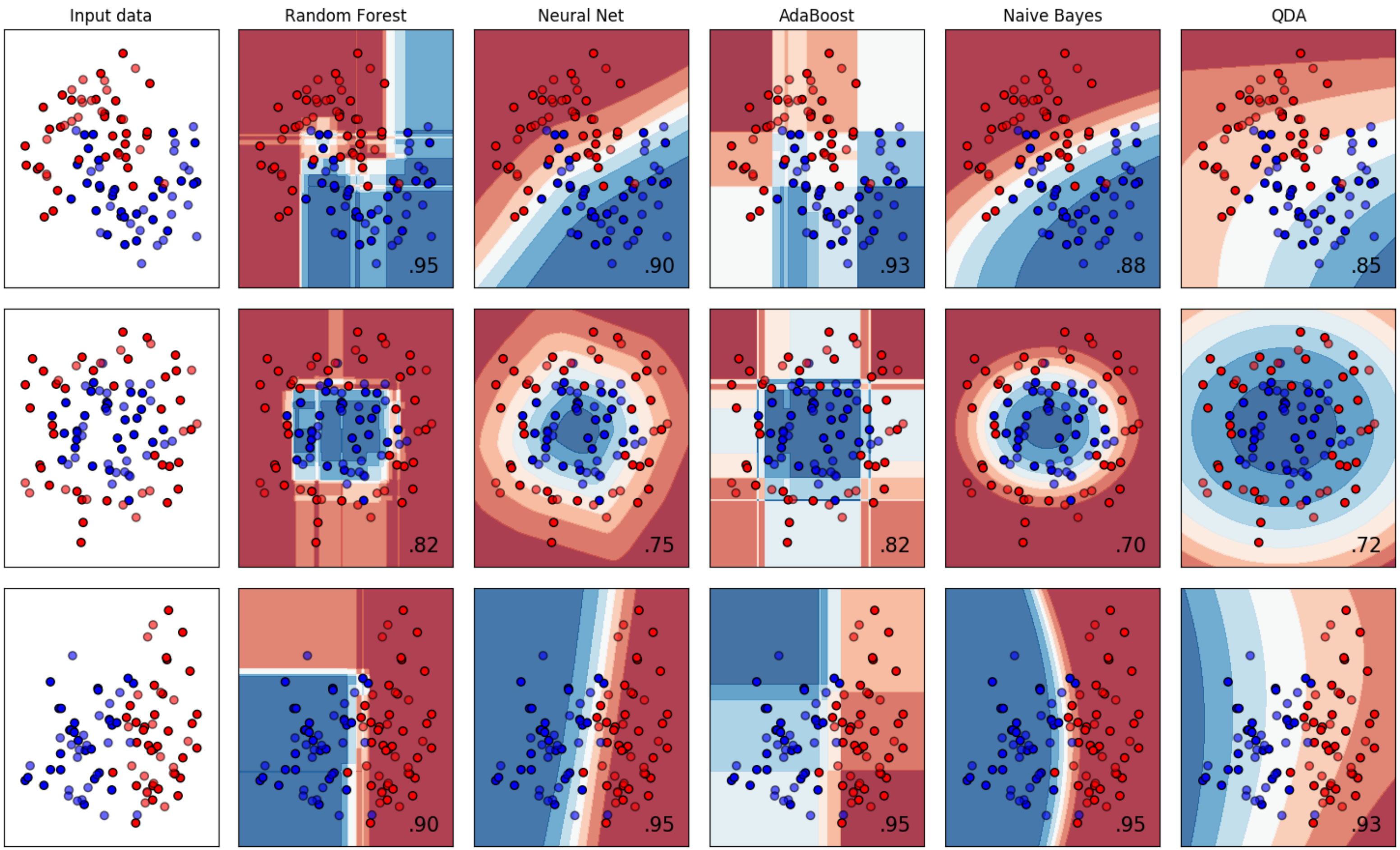
```
import numpy as np
from sklearn.discriminant_analysis import LinearDiscriminantAnalysis
from sklearn.neighbors import KNeighborsClassifier

# Loading the datasets
X = np.loadtxt("./X.dat")
Y = np.loadtxt("./Y.dat")

# LDA
CLF1 = LinearDiscriminantAnalysis()
CLF1.fit(X, Y)
LDA_Score = CLF1.score(X,Y)
print("LDA Score=", LDA_Score)

# KNN
CLF2 = KNeighborsClassifier(n_neighbors=3)
CLF2.fit(X, Y)
KNN_Score = CLF2.score(X,Y)
print("KNN Score=", KNN_Score)
```

Metrics & Validation

Comparison of Classifiers - 1



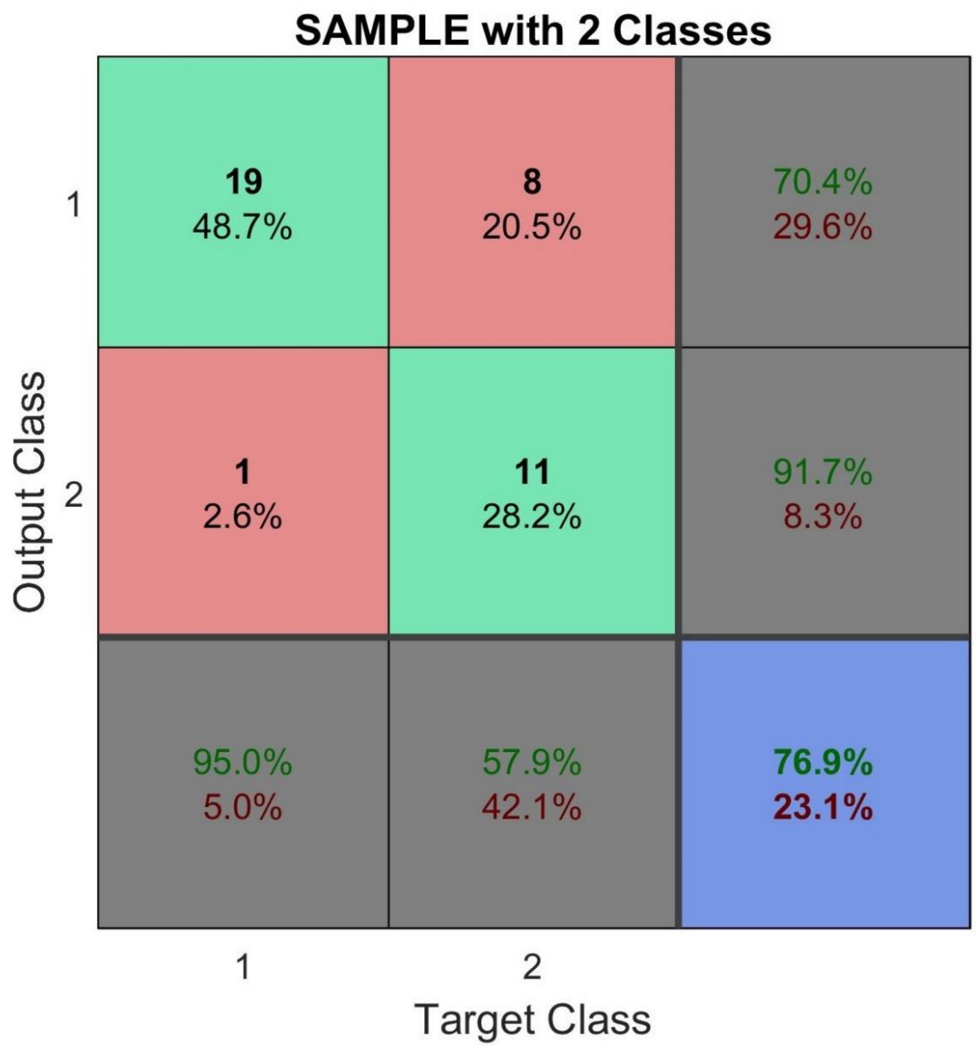
Comparison of Classifiers - 2



Confusion Matrix

		Actual Class	
		P'	N'
Predicted Class	P	True Positive (TP)	False Positive (FP)
	N	False Negative (FN)	True Negative (TN)

Confusion Matrix (Example)



ROC
(receiver operating characteristic)

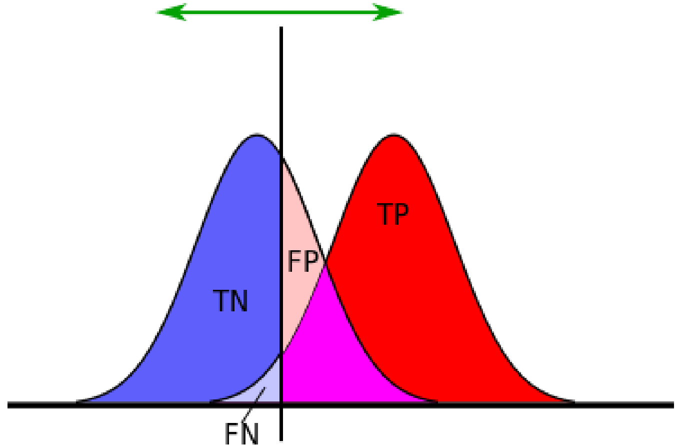
- Plot true positive fraction
  = (P (classified + | actual +)
  = Sensitivity

Vs

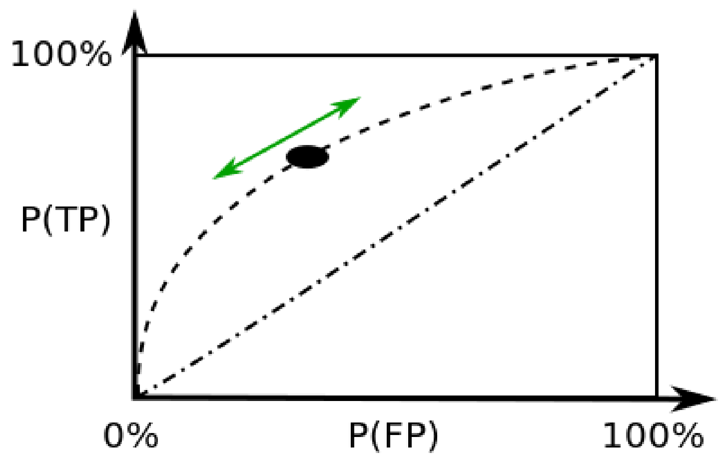
- False positive fraction
  = (P (classified + | actual -)
  = 1 - specificity

Each point on the **ROC** curve represents a sensitivity/specificity pair corresponding to a particular decision threshold.

Threshold



TP	FP
FN	TN

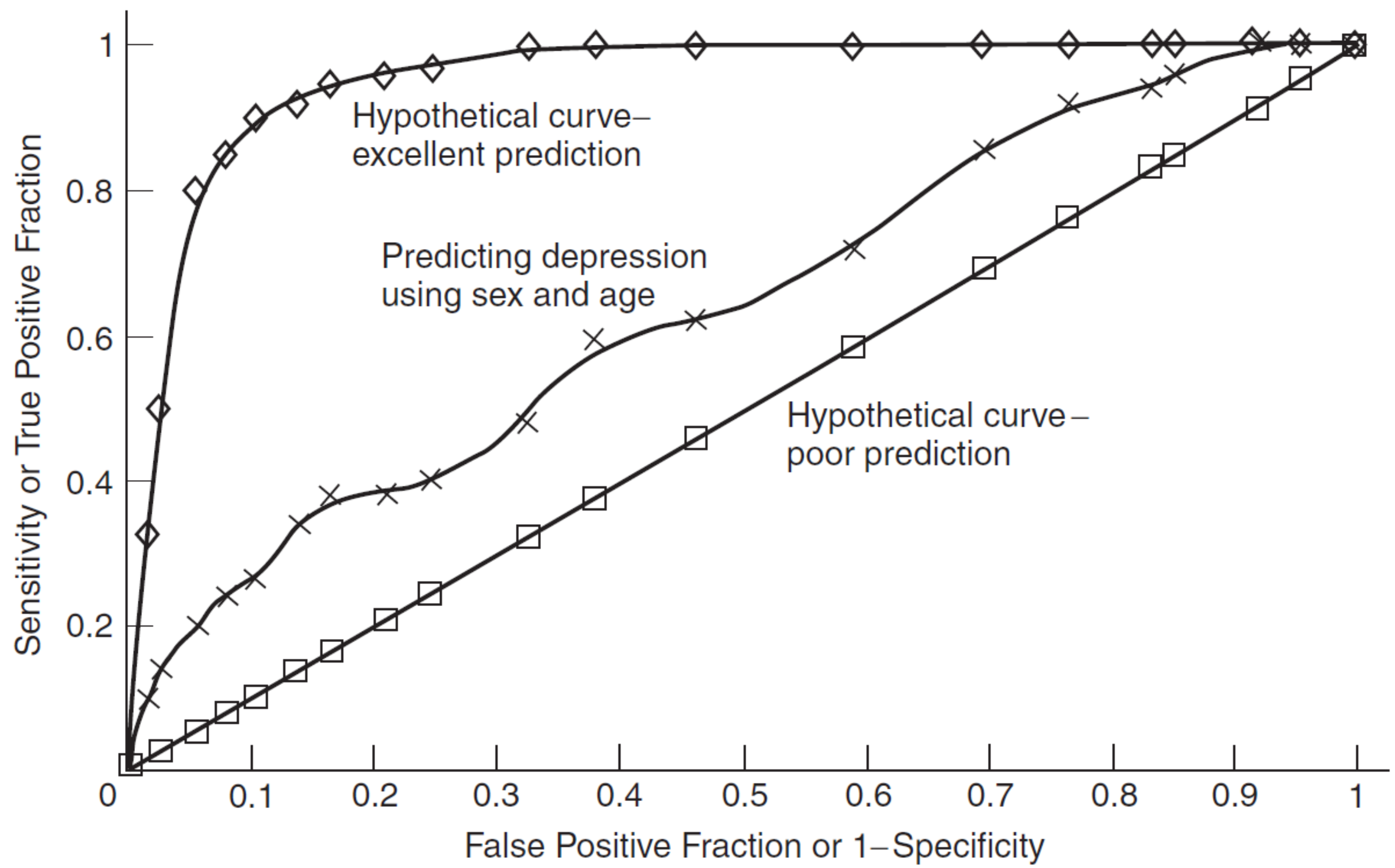


- ↓ threshold ➔ FP ↑ & FN ↓ & TP ↑ & TN ↓)

ROC Curve (p. 297)

FIGURE 12.6

ROC Curve from Logistic Regression for the Depression Data Set



ROC Curve

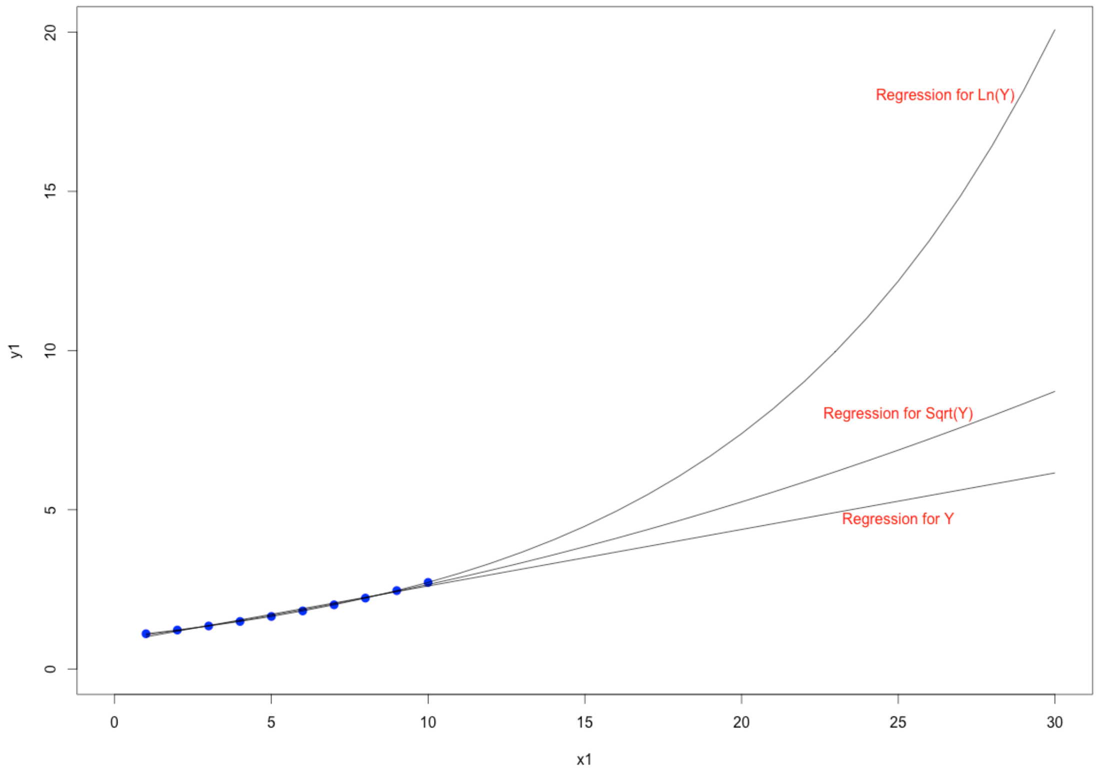
- Area under the curve = a measure of how good the model is. It's between 0.5 and 1.0
- Area under the curve = probability that a randomly chosen positive subject is correctly classified with greater likelihood as positive than a randomly chosen negative subject
- Roughly: Area under the curve is the probability of correct classification

Summary of Classification Overfitting

Given the following data sample:

X	1	2	3	4	5	6	7	8	9	10
Y	1.105	1.221	1.350	1.492	1.649	1.822	2.014	2.226	2.460	2.718

Create a linear regression model for Y, LN(Y), and SQRT (Y)
with X from 1 – 30 and graph.
Show R Squared for each.
Regression for Y:
R-square for Y: 0.9843772
Regression for sqrt(Y)
R-square for sqrt(Y): 0.996024
Regression for ln(Y)
R-square for ln(Y): 1



Cross Validation

1. Create a K-Fold Partition of your Dataset D.
2. For each of K experiments, use K-1 Folds for training and the remaining Fold for testing. For example, if K = 4:

Experiment 1	F1	F2	F3	F4
Experiment 2	F1	F2	F3	F4
Experiment 3	F1	F2	F3	F4
Experiment 4	F1	F2	F3	F4

- **True error rate estimate is: (1/K)Σ (E_i) = Mean of individual error rates.**
- **Available in SAS - PROC GLMSELECT;**
- **Available in several libraries in R, e.g. “boot” “DAAG” “Design”.**

STEVENS
INSTITUTE of TECHNOLOGY
School of Business

stevens.edu

Amir H Gandomi; PhD
Assistant Professor of Analytics & Information Systems
a.h.gandomi@stevens.edu