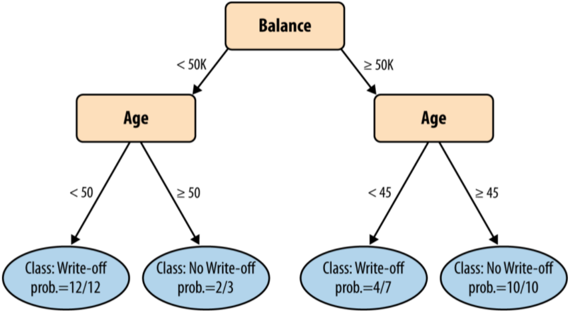
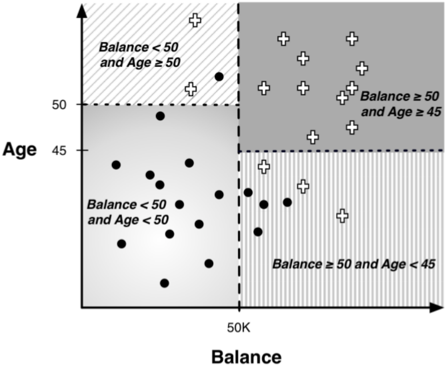
Fundamental Methods of Data Science

Class 7

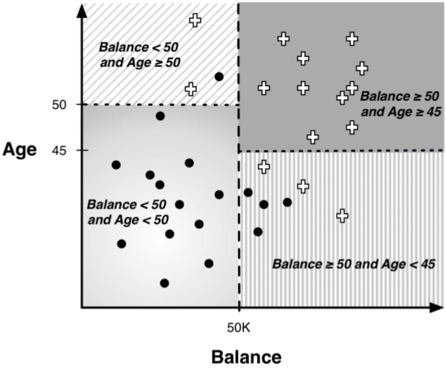
Tree Classification



Tree Classification

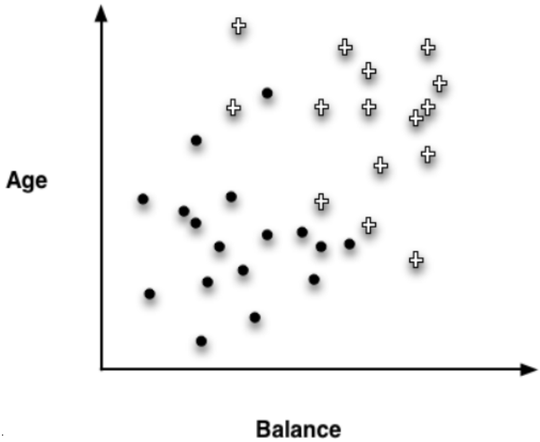


Tree Classification



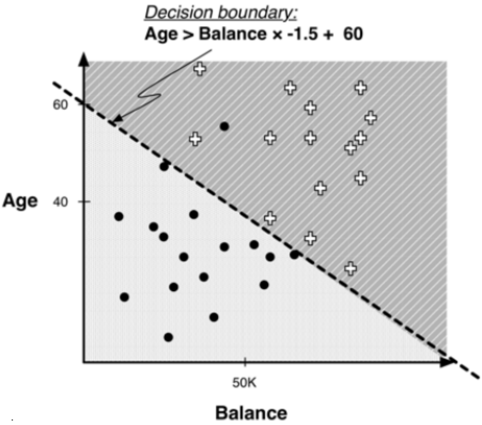
- We can continue classification
  - What is the problem with that?

Classification

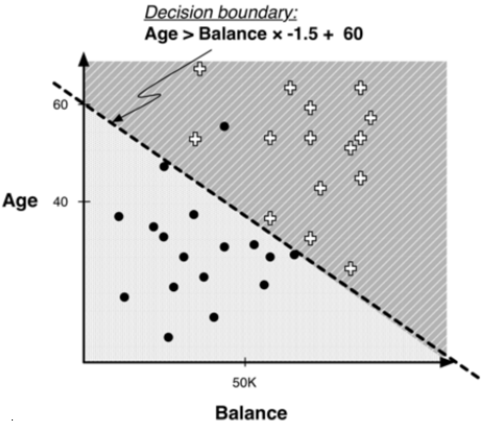


- Can we do better?

Linear Classifiers

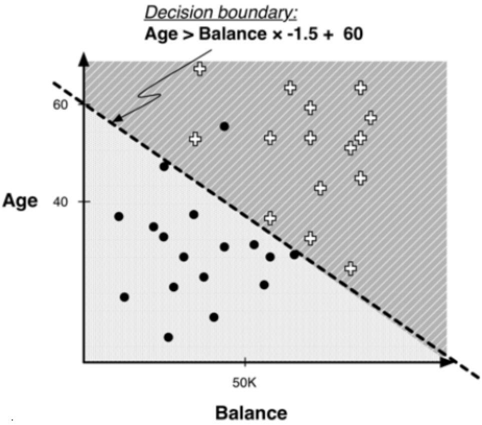


Linear Classifiers



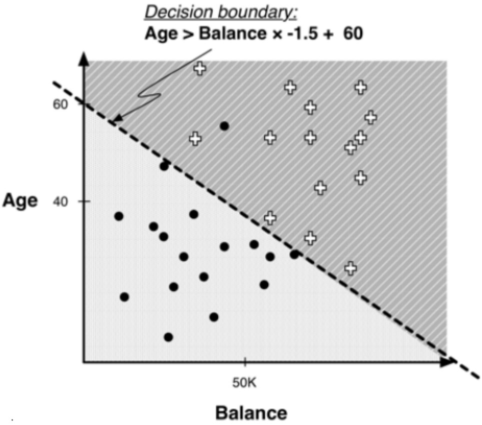
- Line is denoted by the linear equation
  - $Age = (-1.5) \times Balance + 60$

Linear Classifiers



- Line is denoted by the linear equation
  - $Age = (-1.5) \times Balance + 60$
- How can we use it for classification?

Linear Discriminant Functions



- Linear discriminant

$$class(x) = \begin{cases} + \text{ if } 1.0 \times Age \; - 1.5 \times Balance + 60 > 0 \\ \bullet \text{ if } 1.0 \times Age \; - 1.5 \times Balance + 60 \leq 0 \end{cases}$$

Linear Discriminant Functions

$$class(x) = \begin{cases} + \text{ if } 1.0 \times Age - 1.5 \times Balance + 60 > 0 \\ \bullet \text{ if } 1.0 \times Age - 1.5 \times Balance + 60 \leq 0 \end{cases}$$

- How can we obtain such a model?

Linear Discriminant Functions

$$class(x) = \begin{cases} + \text{ if } 1.0 \times Age \; - 1.5 \times Balance + 60 > 0 \\ \bullet \text{ if } 1.0 \times Age \; - 1.5 \times Balance + 60 \leq 0 \end{cases}$$

- How can we obtain such a model?
- $A \times Age + B \times Balance + C$
  - Use data to learn the values of A, B and C

Linear Discriminant Functions

$$class(x) = \begin{cases} + \text{ if } 1.0 \times Age \ - 1.5 \times Balance + 60 > 0 \\ \bullet \text{ if } 1.0 \times Age \ - 1.5 \times Balance + 60 \leq 0 \end{cases}$$

- How can we obtain such a model?
- $A \times Age + B \times Balance + C$
  - Use data to learn the values of A, B and C
- Can you see another advantage over Classification Trees?

Linear Discriminant Functions

$$class(x) = \begin{cases} + \text{ if } 1.0 \times Age \ - 1.5 \times Balance + 60 > 0 \\ \bullet \text{ if } 1.0 \times Age \ - 1.5 \times Balance + 60 \leq 0 \end{cases}$$

- How can we obtain such a model?

- $A \times Age + B \times Balance + C$
  - Use data to learn the values of A, B and C
- Can you see another advantage over Classification Trees?
  - We get an actual value for free!
  - $f(x) = x['Age'] - 1.5 * x['Balance'] + 60$

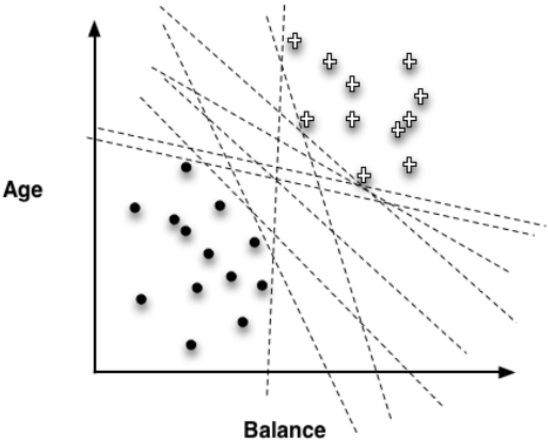
Linear Discriminant Functions vs Classification Trees

- Classification Trees
  - Classification models
  - Use IG to choose features
  - Induct a model
- Discriminant Functions
  - Mathematical formulae
  - Build a model (still need to know which features to use)
  - Tune it according to data

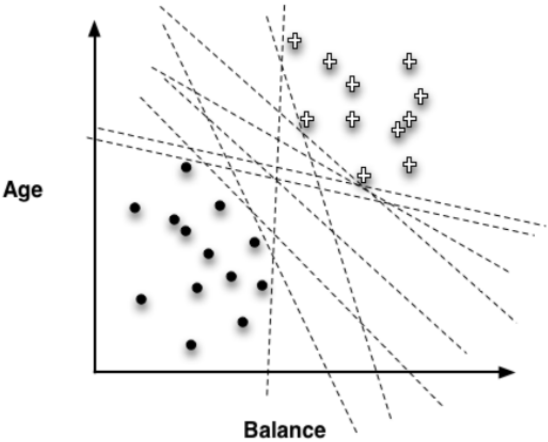
Linear Discriminant Functions and Classification

- Given such a function
  - $f(x) = x['Age'] - 1.5 * x['Balance'] + 60$
- Use the line for classification
  - Positive (Above the line)
  - Negative (Below the line)
- Can be extended to more than two features

Possible Models



Possible Models



- Which one to choose?

Objective Functions

- "Best" line depends on the objective function
  - Objective function should represent our goal

Objective Functions

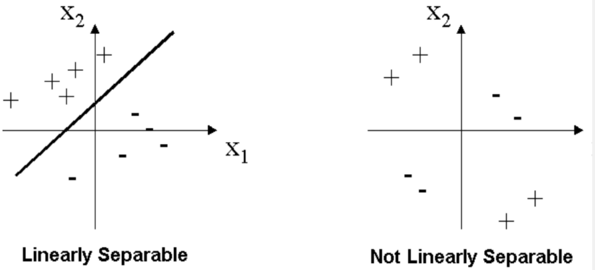
- "Best" line depends on the objective function
  - Objective function should represent our goal
- What about instances misclassified by the model?

Objective Functions

- "Best" line depends on the objective function
  - Objective function should represent our goal
- What about instances misclassified by the model?
  - We can penalize those

Perceptron - A Simple Linear Discriminant Function Learner

- We will see an algorithm for computing such a function in case the data is linearly separable



Perceptron

- We want to learn a function
  - $w_1 \cdot x + w_2 \cdot y + w_0 \cdot 1 = 0$
- By using an instance vector
  - $[(x_1, y_1), \ldots, (x_n, y_n)]$

Perceptron

- We want to learn a function
  - $w_1 \cdot x + w_2 \cdot y + w_0 \cdot 1 = 0$
- By using an instance vector
  - $[(x_1, y_1), \ldots, (x_n, y_n)]$
- We start by arbitrary weights w_0, w_1, w_2
- We adjust them each time they fail to properly classify a point

Perceptron Learning Algorithm

$$\text{Desired output} \qquad d(n) = \begin{cases} +1 & \text{if } x(n) \in \text{set } A \\ -1 & \text{if } x(n) \in \text{set } B \end{cases}$$

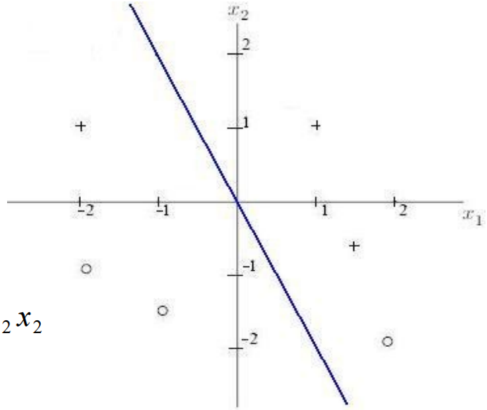
1. Select a random instance n
2. If $d(n)$ is correct, do nothing
3. Else, modify the weights
   - $w_i = w_i + \mu d(n) x_i(n)$
   - μ is the learning rate which must be small in order to avoid misplacing the classifier
4. Repeat until all instances are classified correctly

Example

Initial Values:

$\eta = 0.2$

$$w = \begin{pmatrix} 0 \\ 1 \\ 0.5 \end{pmatrix}$$



$$0 = w_0 + w_1 x_1 + w_2 x_2$$
$$= 0 + x_1 + 0.5x_2$$
$$\Rightarrow \quad x_2 = -2x_1$$

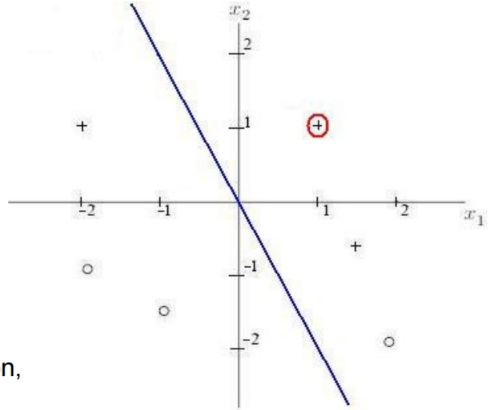
Example

$\eta = 0.2$

$$w = \begin{pmatrix} 0 \\ 1 \\ 0.5 \end{pmatrix}$$

$x_1 = 1,\ x_2 = 1$
$w^T x > 0$

Correct classification,
no action



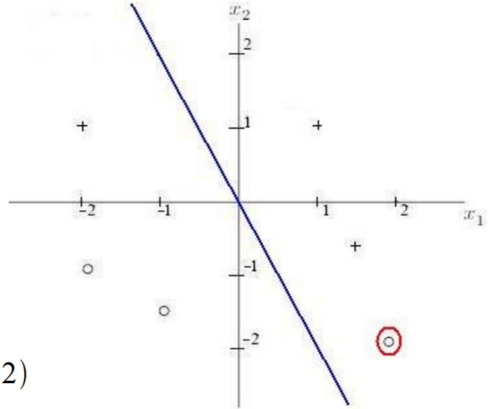
Example

$\eta = 0.2$

$$w = \begin{pmatrix} 0 \\ 1 \\ 0.5 \end{pmatrix}$$

$x_1 = 2, x_2 = -2$

$$w_0 = w_0 - 0.2 * 1$$
$$w_1 = w_1 - 0.2 * 2$$
$$w_2 = w_2 - 0.2 * (-2)$$



Example

$\eta = 0.2$

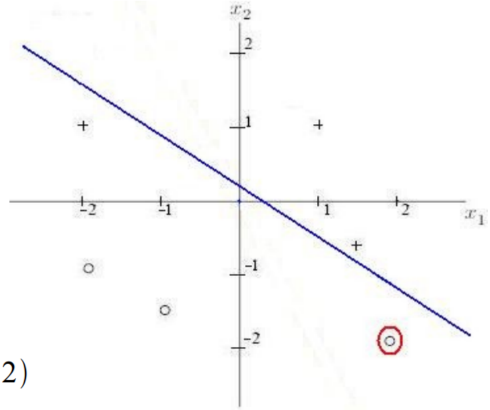
$$w = \begin{pmatrix} -0.2 \\ 0.6 \\ 0.9 \end{pmatrix}$$

$x_1 = 2, x_2 = -2$

$$w_0 = w_0 - 0.2 * 1$$

$$w_1 = w_1 - 0.2 * 2$$

$$w_2 = w_2 - 0.2 * (-2)$$



Example

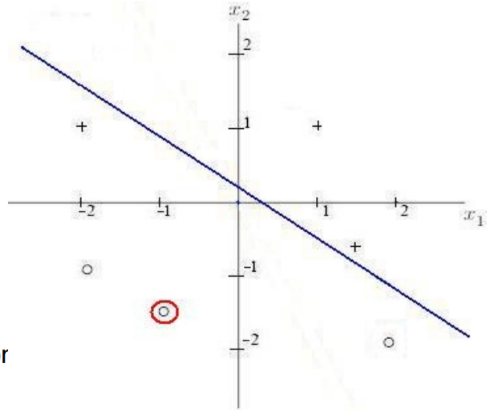
$\eta = 0.2$

$$w = \begin{pmatrix} -0.2 \\ 0.6 \\ 0.9 \end{pmatrix}$$

$x_1 = -1,\ x_2 = -1.5$

$w^T x < 0$

Correct classificatior
no action



Example

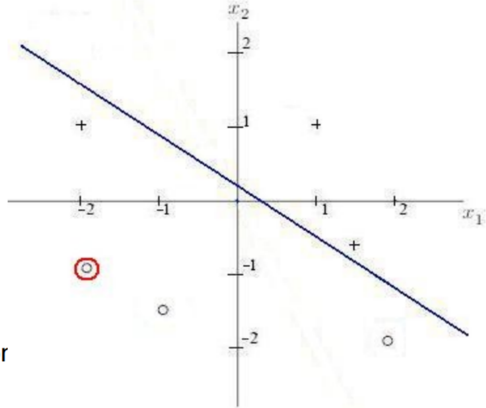
$\eta = 0.2$

$$w = \begin{pmatrix} -0.2 \\ 0.6 \\ 0.9 \end{pmatrix}$$

$x_1 = -2,\ x_2 = -1$

$w^T x < 0$

Correct classificatior
no action



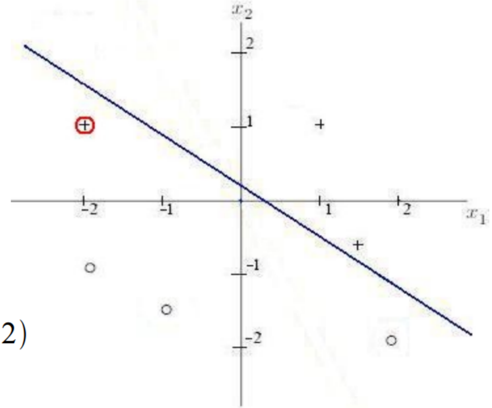
Example

$\eta = 0.2$

$$w = \begin{pmatrix} -0.2 \\ 0.6 \\ 0.9 \end{pmatrix}$$

$x_1 = -2,\ x_2 = 1$

$$w_0 = w_0 + 0.2 * 1$$
$$w_1 = w_1 + 0.2 * (-2)$$
$$w_2 = w_2 + 0.2 * 1$$



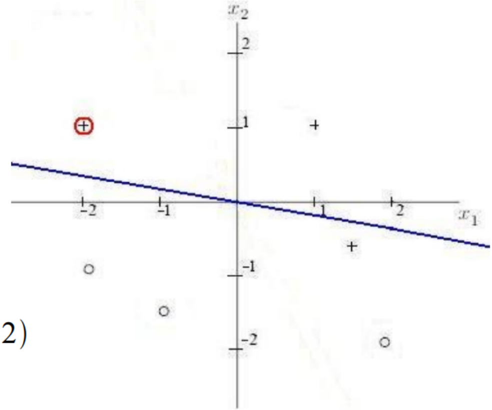
Example

$\eta = 0.2$

$$w = \begin{pmatrix} 0 \\ 0.2 \\ 1.1 \end{pmatrix}$$

$x_1 = -2,\ x_2 = 1$

$$w_0 = w_0 + 0.2 * 1$$
$$w_1 = w_1 + 0.2 * (-2)$$
$$w_2 = w_2 + 0.2 * 1$$



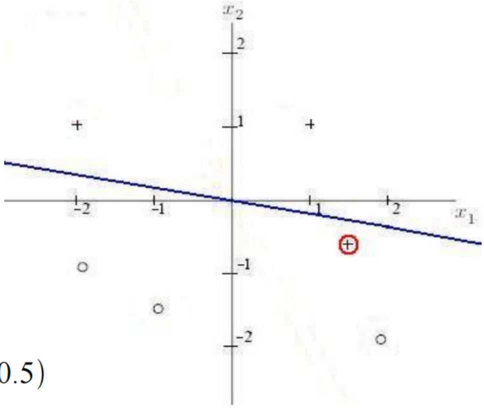
Example

$\eta = 0.2$

$$w = \begin{pmatrix} 0 \\ 0.2 \\ 1.1 \end{pmatrix}$$

$x_1 = 1.5,\ x_2 = -0.5$

$$w_0 = w_0 + 0.2 * 1$$
$$w_1 = w_1 + 0.2 * 1.5$$
$$w_2 = w_2 + 0.2 * (-0.5)$$



Example

$\eta = 0.2$

$$w = \begin{pmatrix} 0.2 \\ 0.5 \\ 1 \end{pmatrix}$$

$x_1 = 1.5, x_2 = -0.5$

$$w_0 = w_0 + 0.2 * 1$$

$$w_1 = w_1 + 0.2 * 1.5$$

$$w_2 = w_2 + 0.2 * (-0.5)$$