LINEAR CLASSIFIERS

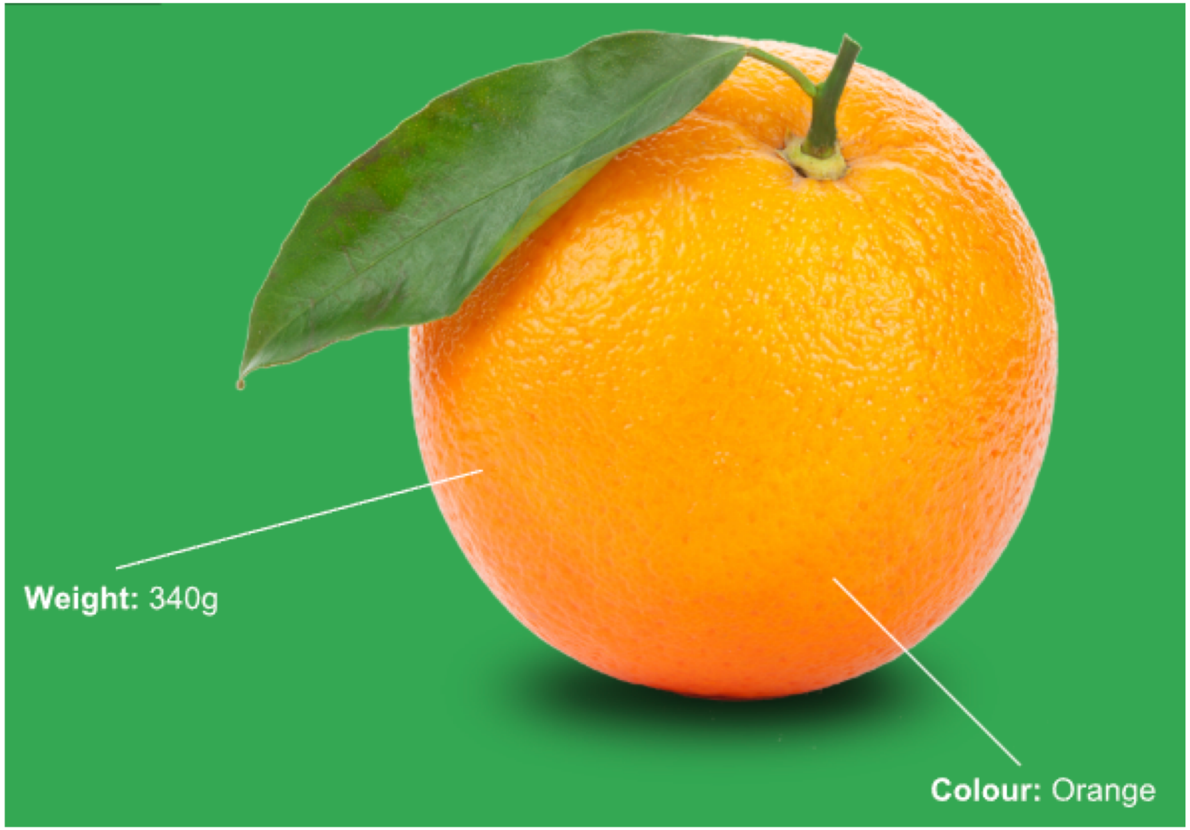
Classifiers

Motivation

- Simple and popular **supervised** models
- Suitable for regression and classification
  (*we'll study regression later*)
- Linear models inspired a lot of other methods like Neural Networks and Support Vector Machines.

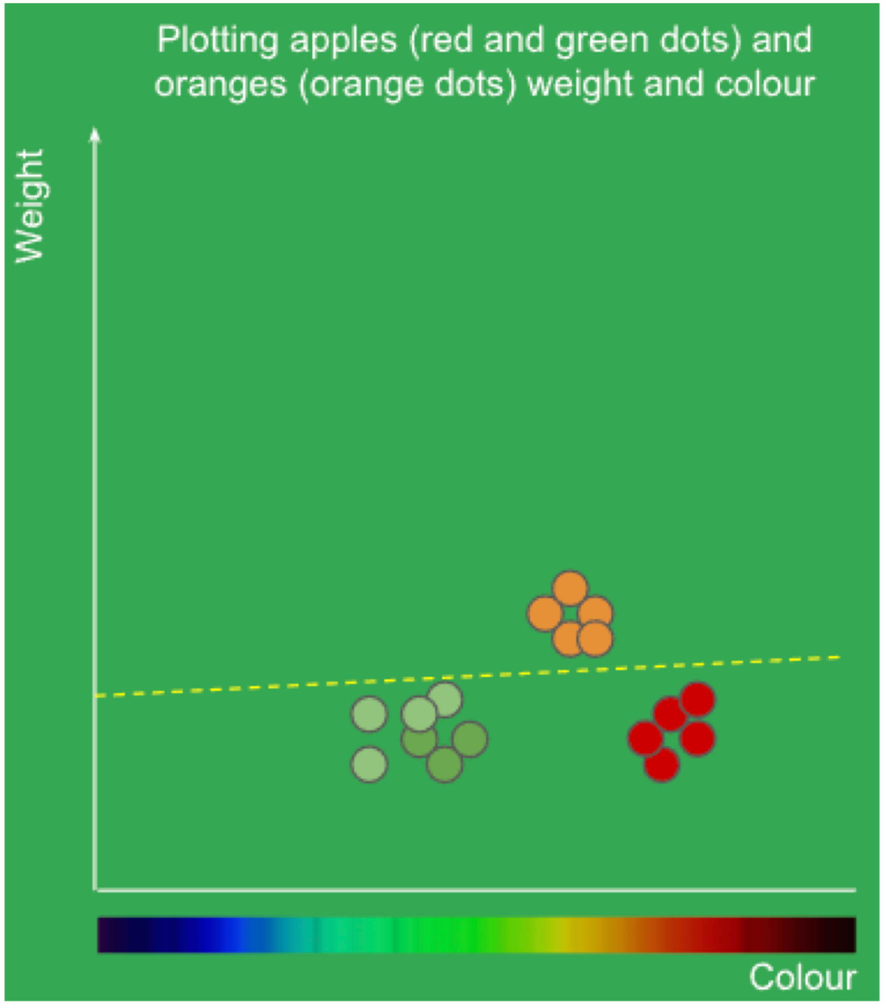
Introduction

Representation space: attributes as coordinates



Introduction

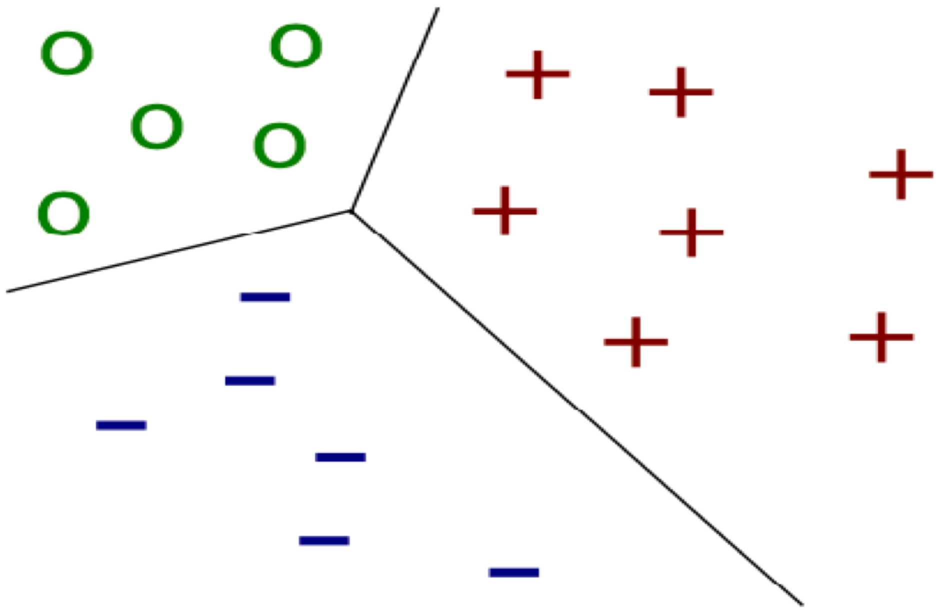
Separate oranges
from red and green
apples
using a **linear** frontier
in the attribute space



Illustrations: Jason Mayes Machine Learning 101

Introduction

Multi-class linear classifier



Introduction

Supervised classification

Weight	Color	Seeds	Fruit	
150	80	8	0	apple
200	112	6	1	orange
170	120	8	1	orange
210	105	7	1	orange
180	130	9	0	apple

attributes x_i class (target) y

Linear model:

$$\sum_i w_i x_i + w_0 = \begin{cases} < 0 \Rightarrow & \text{class } 0 \\ \geq 0 \Rightarrow & \text{class } 1 \end{cases}$$

The decision boundary is an hyperplane

Linear separators

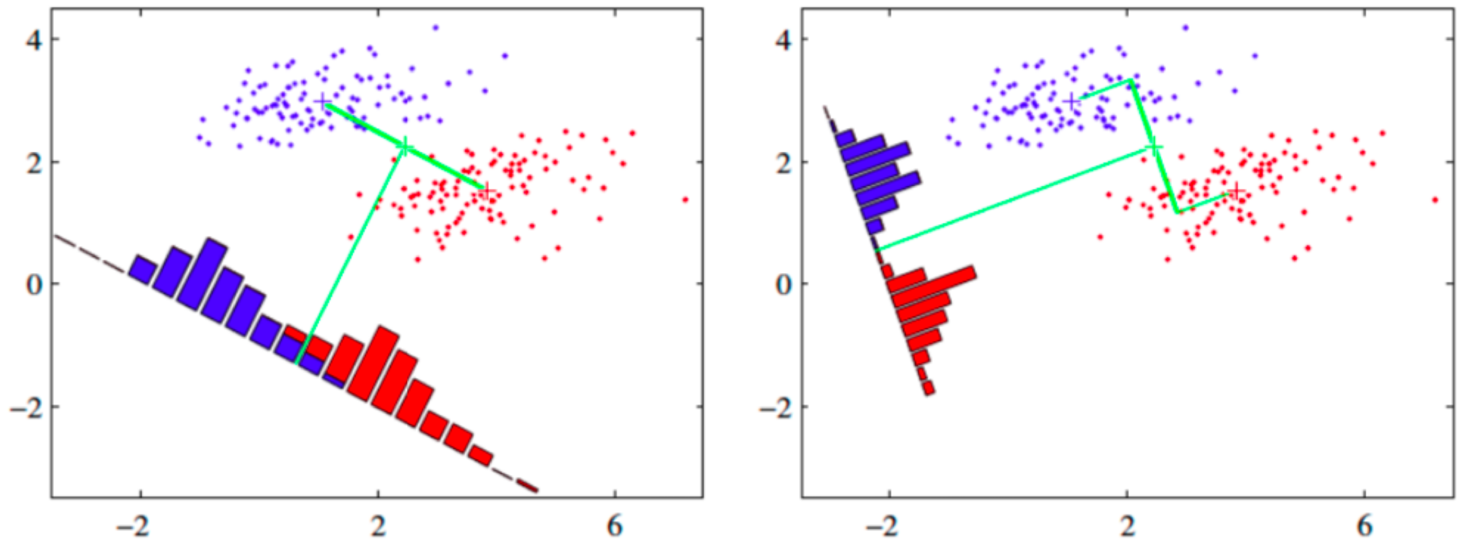
How can we find the separating hyperplane ? w_i

Many approaches:

- MSE (linear regression, studied later)
- **LDA linear discriminant analysis**
- **Logistic Regression** (a classification method)
- Perceptron (studied later)
- Support Vector Machines (studied later)

Linear Discriminant Analysis

Fischer, 1936



Finding the direction maximizing a ratio of “between-class variance” to “within-class variance”

Figure: Cheng Yen Li, 2014

Linear Discriminant Analysis

- Can be generalized for any number of classes
- Similar to PCA, rely on the eigen-decomposition of covariance matrices, leading to high complexity $O(n^3)$

Logistic Regression

… should be called Logistic Discrimination

Basic assumption: the difference between the logarithms of the class-conditional density functions is linear in the variables x:

$$\log\left(\frac{p(\boldsymbol{x}|\omega_1)}{p(\boldsymbol{x}|\omega_2)}\right) = \beta_0 + \boldsymbol{\beta}^T\boldsymbol{x}$$

where ω_1 *,* ω_2 *are the classes*

Logistic Regression

The previous equation can be rewritten as

$$p(\omega_2|\boldsymbol{x}) = \frac{1}{1 + \exp(\beta_0' + \boldsymbol{\beta}^T \boldsymbol{x})}$$

$$p(\omega_1|\boldsymbol{x}) = \frac{\exp(\beta_0' + \boldsymbol{\beta}^T \boldsymbol{x})}{1 + \exp(\beta_0' + \boldsymbol{\beta}^T \boldsymbol{x})}$$

(easy to generalize to arbitrary number of classes)

Logistic Regression

Decision between two classes

$$\text{assign } \boldsymbol{x} \text{ to} \begin{cases} \omega_1 \\ \omega_2 \end{cases} \text{if } \frac{p(\omega_1|\boldsymbol{x})}{p(\omega_2|\boldsymbol{x})} \begin{cases} > \\ < \end{cases} 1$$

which leads to

$$\text{assign } \boldsymbol{x} \text{ to} \begin{cases} \omega_1 \\ \omega_2 \end{cases} \text{if } \beta_0' + \boldsymbol{\beta}^T \boldsymbol{x} \begin{cases} > \\ < \end{cases} 0$$

Again, a linear model.

Logistic Regression

Maximum Likelihood

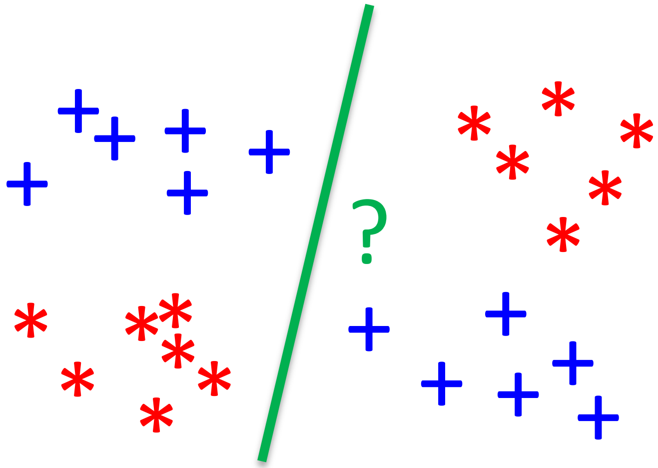
For C classes, the likelihood of the examples is

$$L = \prod_{i=1}^{C} \prod_{r=1}^{n_i} p(\boldsymbol{x}_{ir} | \omega_i)$$

Maximizing L is equivalent to maximize L'

$$\log(L') = \sum_{s=1}^{C} \sum_{r=1}^{n_s} \log(p(\omega_s | \boldsymbol{x}_{sr}))$$

Linear separability ?



Solutions:

- Add or generate attributes
- Use non-linear models

Conclusion

- Linear models are easy to build and interpret

- Logistic regression is especially popular because it offer good performances and its outputs are interpretable as probabilities.

References

Books

- T. Hastie, R. Tibshirani, J. Friedman. The Elements of Statistical Learning. Springer, 2017 https://web.stanford.edu/~hastie/ElemStatLearn/
- Andrew Webb, K. Copsey. Statistical Pattern Recognition 3rd Edition. Wiley, 2011.

Tutorials

- R. Shaw. Logistic Regression: A Concise Technical Overview. https://www.kdnuggets.com/2018/02/logistic-regression-concise-technical-overview.html
- Cheng Li, Bingyu Wang Fisher Linear Discriminant Analysis, 2014 https://www.semanticscholar.org/paper/Fisher-Linear-Discriminant-Analysis-Li/1ab8ea71fbef3b55b69e142897fadf43b3269463
- Scikit-learn:
  - Generalized Linear Models https://scikit-learn.org/stable/modules/linear_model.html
  - https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html
  - Logistic Regression 3-class Classifier. https://scikit-learn.org/stable/auto_examples/linear_model/plot_iris_logistic.html