INGENIERIA INDUSTRIAL
UNIVERSIDAD DE CHILE

IN4402: Aplicaciones de Probabilidades y Estadística

Machine Learning Methods

ANDRÉS FERNÁNDEZ

BBC | Sign in | Home | News | Sport | Reel | Worklife | Travel | Future | Culture | More | Search

NEWS

Home | Coronavirus | Video | World | US & Canada | UK | Business | Tech | Science | Stories | Entertainment & Arts | More

Tech

Google achieves AI 'breakthrough' by beating Go champion

27 January 2016

Top Stories

Netanyahu out as new Israeli government approved

Benjamin Netanyahu loses his 12-year hold on power as a new Israeli government is approved.

3 hours ago

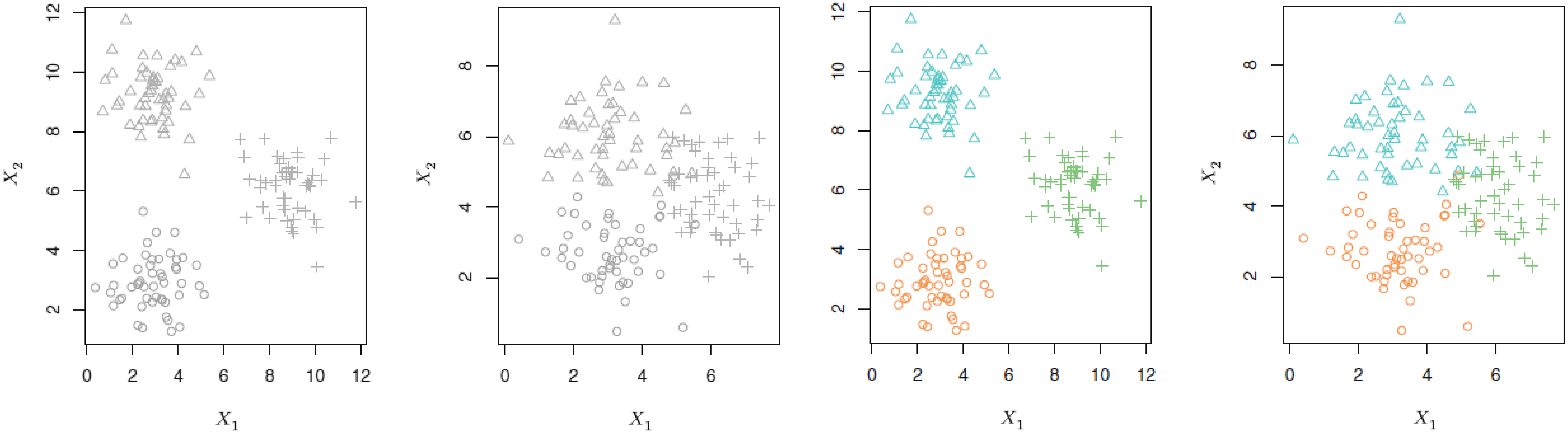
TECHNOLOGY — The New York Times

How Many Computers to Identify a Cat? 16,000

- Statistics draws population **inferences** from a **sample**, and machine learning finds **generalizable predictive patterns** (Bzdok, Altman & Krzywinski, 2018)

- We'll understand *learning* as "generalizing from experiences"
    - We'll feed data-centered experiences to the machine
    - We'll set an <u>*algorithm*</u> so the machine can use the data
    - We'll set a set of <u>*parameters,*</u> so the machine knows how to contrasts the experiences
    - We'll set some <u>*performance*</u> <u>*measures,*</u> so we know if the machine is improving
    - We'll generalize the learning procedure to new data in order to use the patterns learned

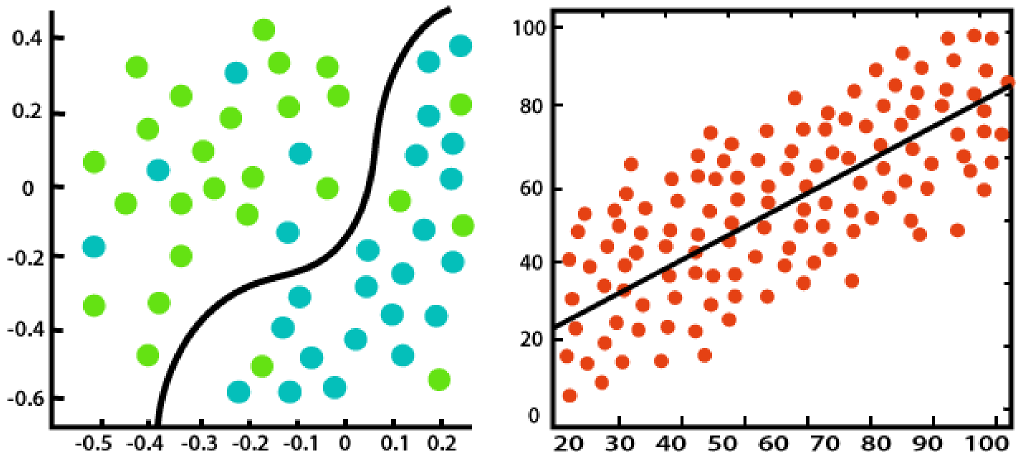
- When using ML approaches, there is a wide range of uses
- By the structure of the algorithm
  - Supervised (we input a model knowing the response)
  - Unsupervised (we ask the model to show inner patterns)



Source: James, Witten, Hastie & Tibshirani (2013) An Introduction to Statistical Learning: with applications in R. New York: Springer

MACHINE LEARNING
INTRODUCTION AND MAIN CONCEPTS

- When using ML approaches, there is a wide range of uses
- By the prediction variable's nature
  - Classification (would this subject be treated/control?)
  - Regression (how much will this subject spend?)



Classification Regression

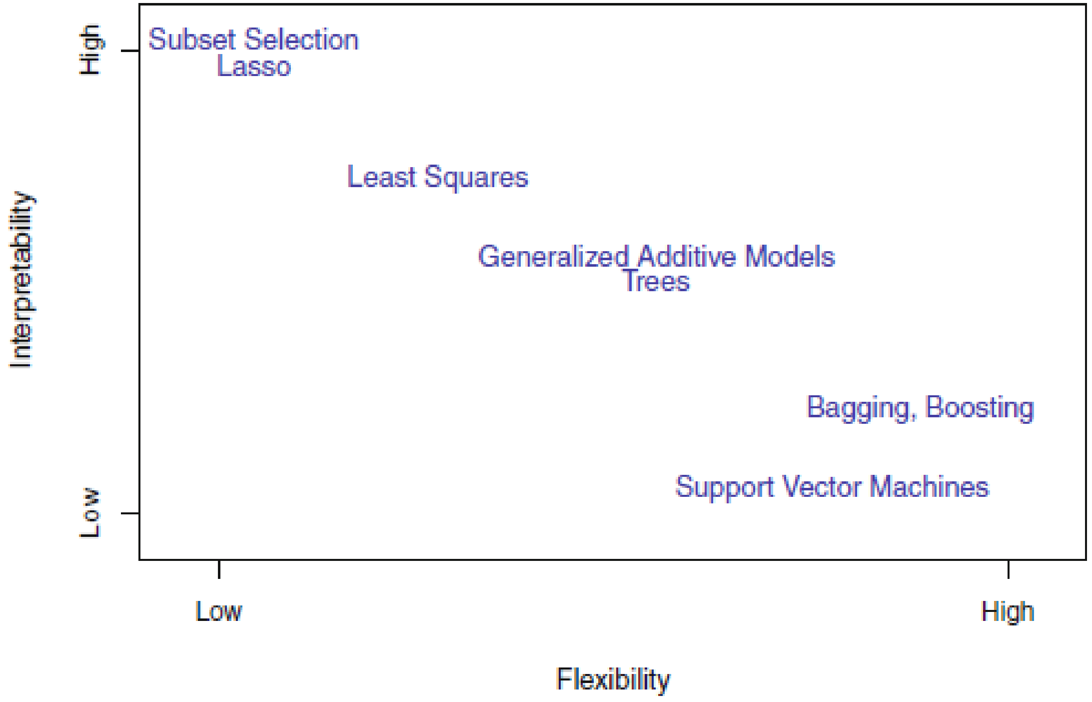
- When using ML approaches, there is a wide range of uses because we want to replicate the data-production process or function according to its patterns

- Which method is the best one? **It depends**
  - Some predict better, other classify better, other work better (un)supervised, etc.
  - Sometimes we can ensamble methods to work optimally

- To assess the models' **performance**:
  - Evaluate the model fit in training samples
  - Evaluate the model's prediction in test samples

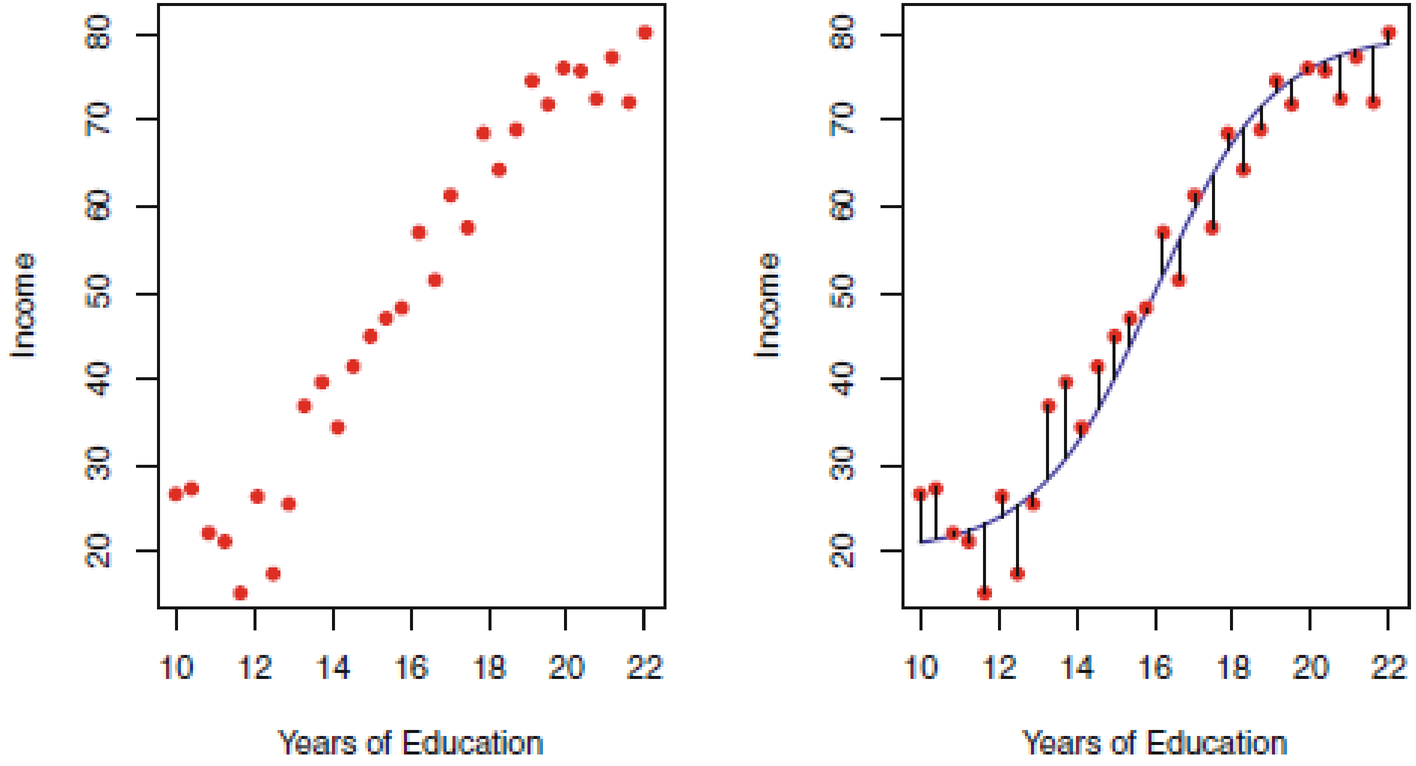
- What happens with **interpretation**?
  - Trade-off between flexibility and interpretation

- There is a trade-off between **interpretation** and **flexibility**
  - In order to understand variable relations we assume simpler forms (i.e. linear models)
  - More flexible models tend to be highly **data-centered** and are more difficult to understand



Source: James, Witten, Hastie & Tibshirani (2013) An Introduction to Statistical Learning: with applications in R. New York: Springer

- What happens with **interpretation**?
- What happens with **generalization**?



Source: James, Witten, Hastie & Tibshirani (2013) An Introduction to Statistical Learning: with applications in R. New York: Springer

INGENIERIA INDUSTRIAL
UNIVERSIDAD DE CHILE

IN4402: Aplicaciones de Probabilidades y Estadística

Machine Learning Performance Evaluation

ANDRÉS FERNÁNDEZ

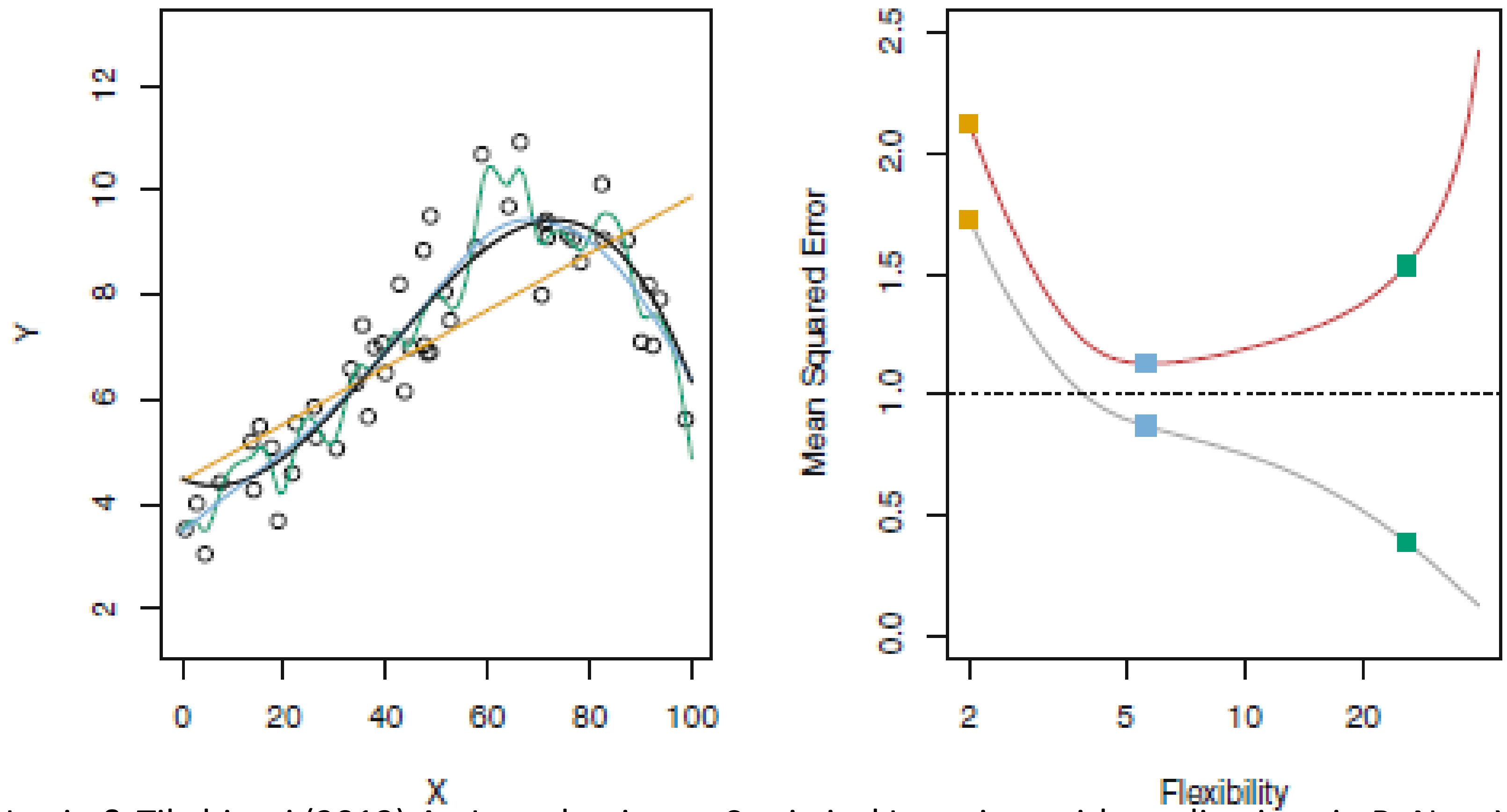
- We want models to effectively predict data using patterns inside the sample
  - How to evaluate better models for prediction purposes?
  - Let’s start with a common measure: *Mean Squared Errors (MSE)*

$$MSE = \left(\frac{1}{n}\right)\sum_{i=1}^{n}\left(y_i - \hat{f}(x_i)\right)^2$$

  - *We don’t need good predictions on known data…*
    - We split in ***training*** and ***test*** samples

Source: James, Witten, Hastie & Tibshirani (2013) An Introduction to Statistical Learning: with applications in R. New York: Springer

- Train vs. test data → training MSE and test MSE
- Overfitting vs. Underfitting



Source: James, Witten, Hastie & Tibshirani (2013) An Introduction to Statistical Learning: with applications in R. New York: Springer

- **Variance vs. Bias**
- The expected value of MSE can be decomposed into three terms:

$$E\left(y_i - \hat{f}(x_i)\right)^2 = Var\left(\hat{f}(x_i)\right) + \left[Bias\left(\hat{f}(x_i)\right)\right]^2 + Var(\varepsilon_i)$$

Source: James, Witten, Hastie & Tibshirani (2013) An Introduction to Statistical Learning: with applications in R. New York: Springer

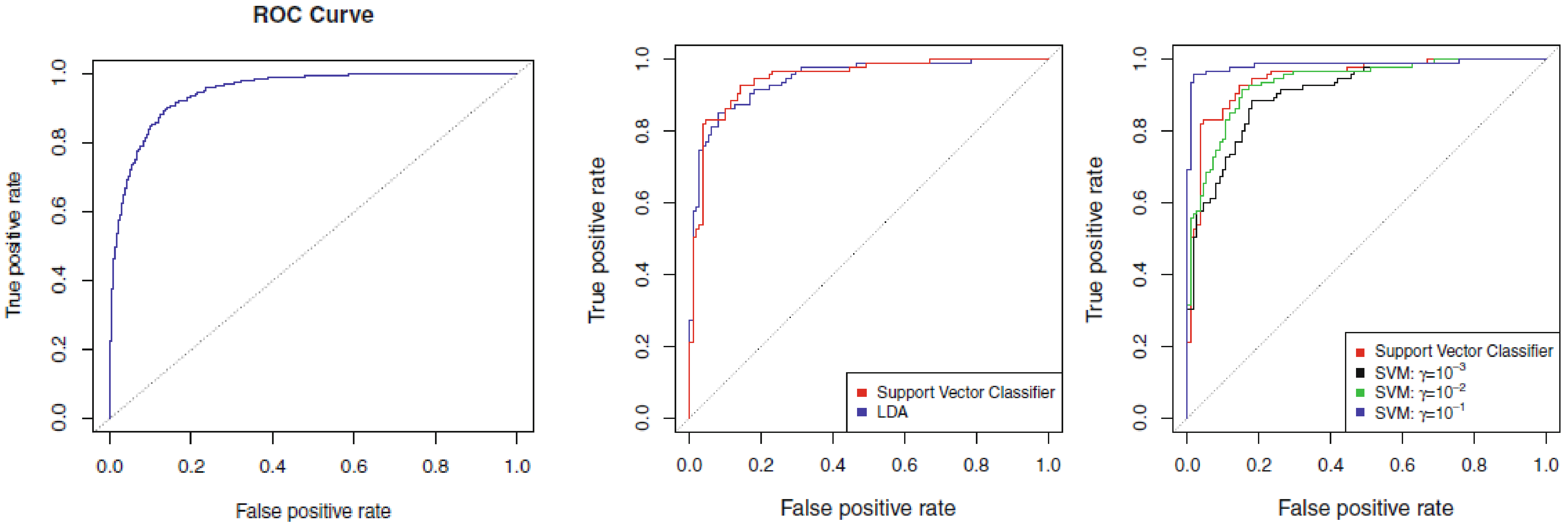
- **Performance evaluation in classification setting:**
- **Error rate:** how many mistakes are made when using a predicting model
- **Contingency tables** of classification:

		Predicted class		
		− or Null	+ or Non-null	Total
True class	− or Null	True Neg. (TN)	False Pos. (FP)	N
	+ or Non-null	False Neg. (FN)	True Pos. (TP)	P
	Total	N^*	P^*	

Name	Definition	Synonyms
False Pos. rate	FP/N	Type I error, 1−Specificity
True Pos. rate	TP/P	1−Type II error, power, sensitivity, recall
Pos. Pred. value	TP/P^*	Precision, 1−false discovery proportion
Neg. Pred. value	TN/N^*	

Source: James, Witten, Hastie & Tibshirani (2013) An Introduction to Statistical Learning: with applications in R. New York: Springer

- **Performance evaluation in classification setting:**
- **ROC:**



Source: James, Witten, Hastie & Tibshirani (2013) An Introduction to Statistical Learning: with applications in R. New York: Springer

- Summary
  - Machine learning methods are widely used for prediction of binary or continuous variables
  - In supervised models, fitting the data and predicting new data is to be balanced
    - This is achieved splitting the data in training and test samples
  - Some indicators, such as MSE and contingency tables are used to assess model predictions
    - MSE must be low both for training and test samples.
    - Contingency tables can be used to build ROC curves to compare models