Supervised vs Unsupervised Learning

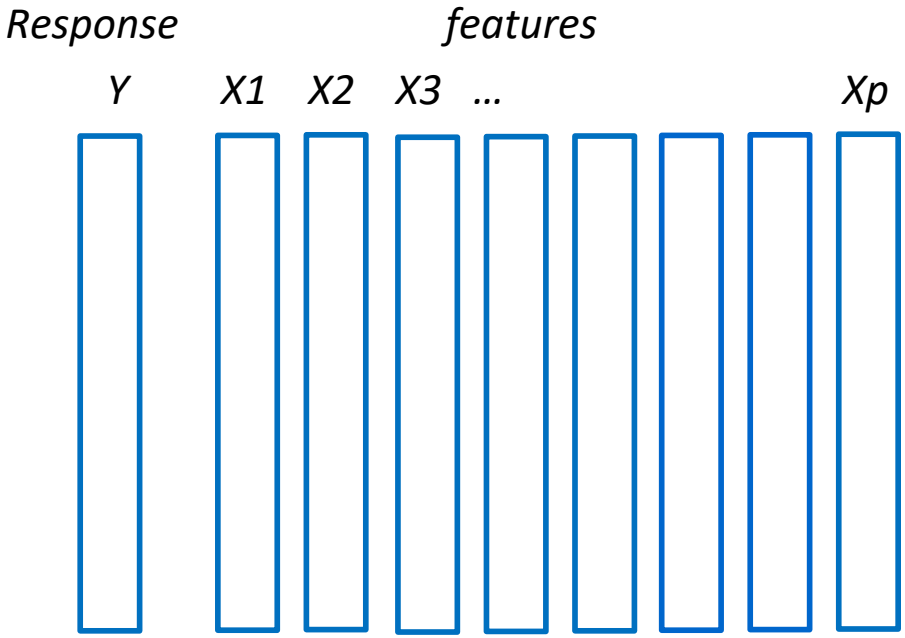
Cesar Acosta Ph.D.

Department of Industrial and Systems Engineering
University of Southern California

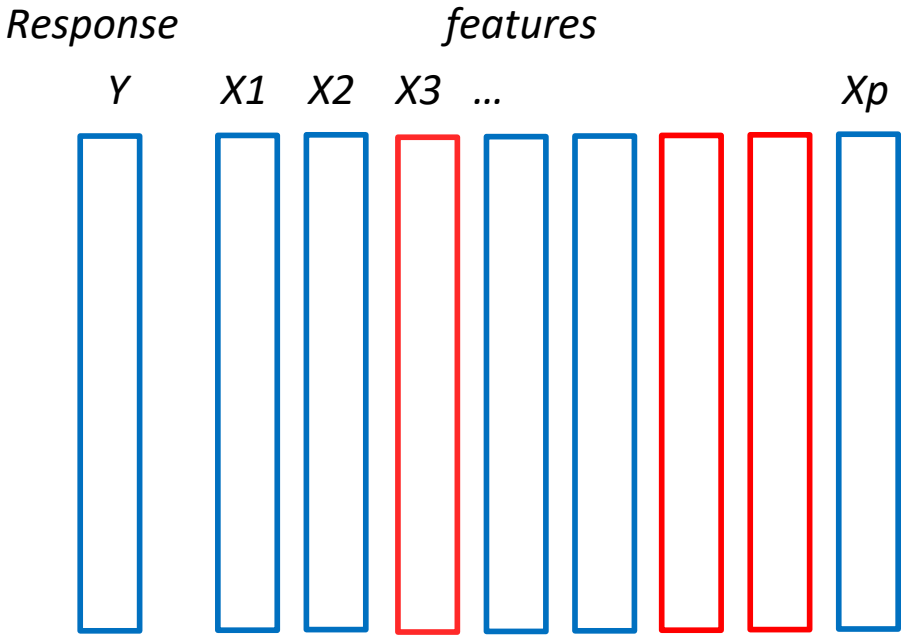
SUPERVISED LEARNING PROBLEM

- *Construct a model relating a response y with predictors x_1 , x_2,…,x_p to*
  - *predict the response y for new observations*
  - *understand the relationship between the response and the predictors*
- *Response and predictors may be numeric or categorical*

Regression



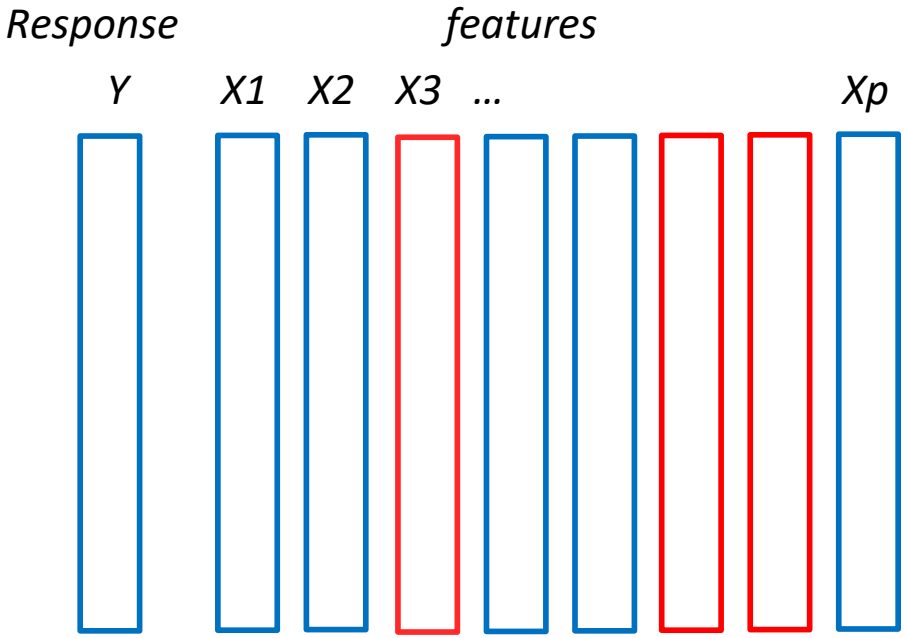
Regression



Blue for numeric predictor

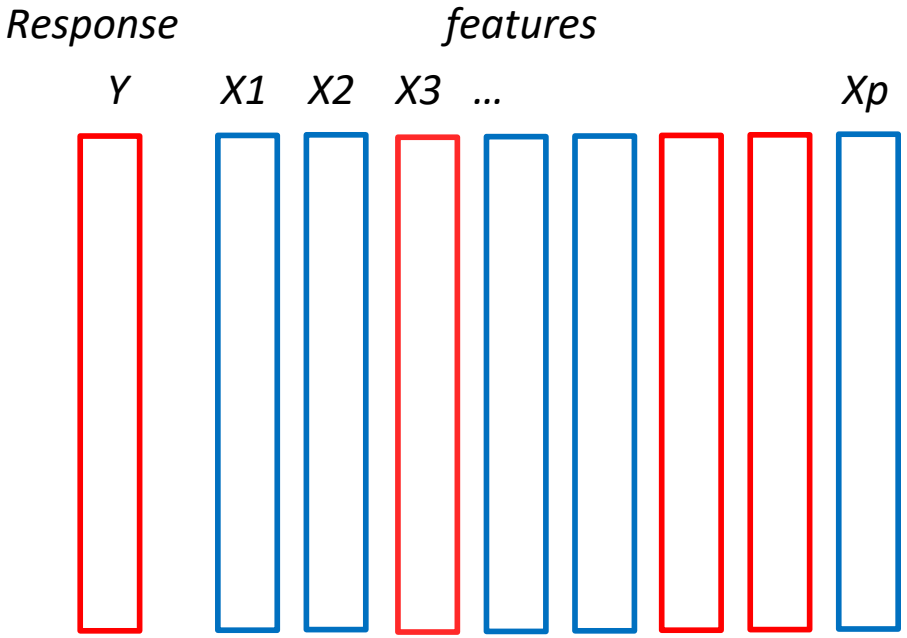
Red for categorical predictor

Regression



Response is numeric in Regression problems

Regression vs Classification



Response is categorical in Classification problems

Linear Regression

- *Response is numeric for linear Regression models*

- *Regression models assume*

  the response is a normal r. variable

Linear Models

Response random variable assumption	*Model*
• *Normal*	• *Linear Regression*
• *Bernoulli*	• *Logistic regression*
• *Binomial*	• *Binomial regression*
• *Negative binomial*	• *Negative binomial regression*
• *Multinomial*	• *Multinomial regression*
• *Poisson*	• *Poisson regression*

Linear Models

Response random variable assumption

- *Normal*
- *Bernoulli*
- *Binomial*
- *Negative binomial*
- *Multinomial*
- *Poisson*

Generalized linear Models (GLM)

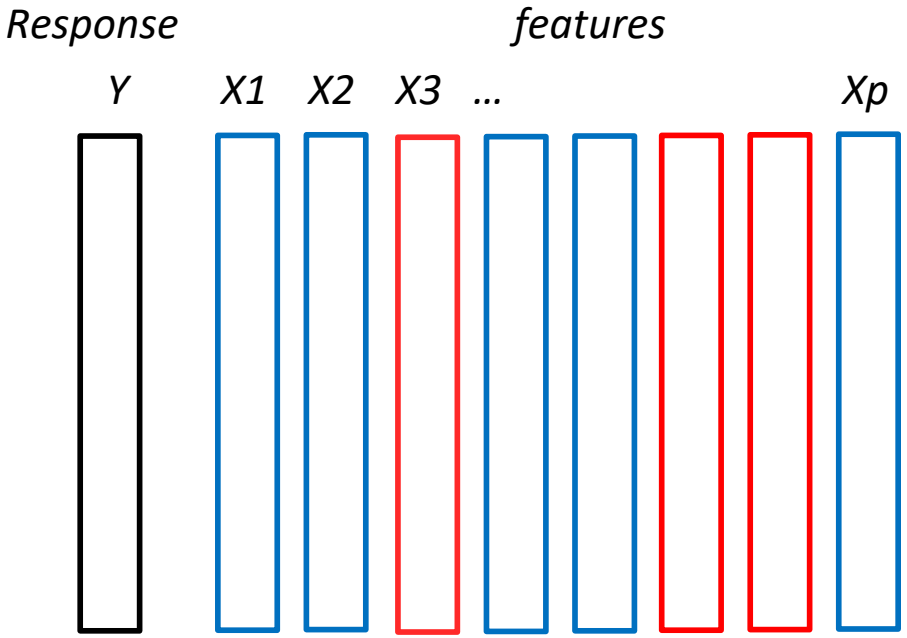
- *Logistic regression*
- *Binomial regression*
- *Negative binomial regression*
- *Multinomial regression*
- *Poisson regression*

UNSUPERVISED LEARNING PROBLEM

- *Observations include features (numerical and categorical) but no associated response*

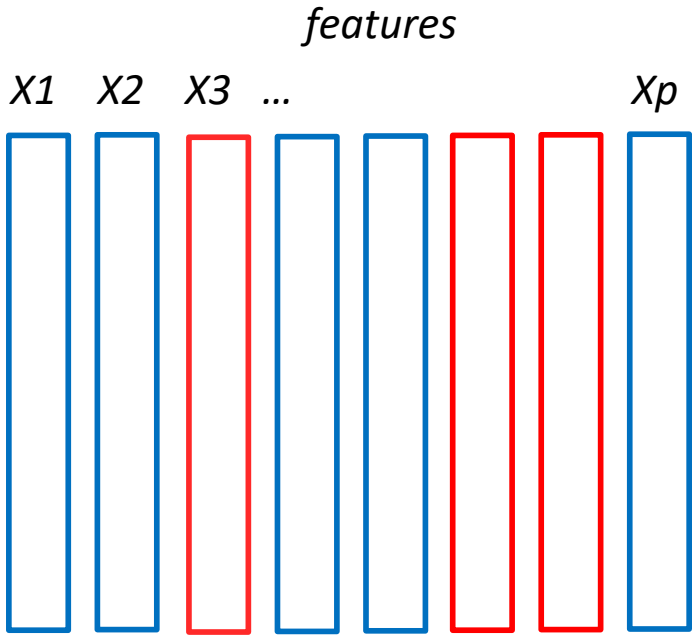
- *Unsupervised since there is no response that can supervise the analysis*

Supervised learning problem



Response may be numeric or categorical

Unsupervised learning



There is no response

UNSUPERVISED LEARNING PROBLEM

- ❑ *There is no response that can supervise the analysis*

Example

- *To determine the performance of a supervised learning model compare predictions with observed Y values*

UNSUPERVISED LEARNING PROBLEM

- ❑ *There is no response that can supervise the analysis*

Example

- *To determine the performance of a supervised learning model compare predictions with observed Y values*
- *This is not possible for unsupervised learning models*
- *Cannot determine the performance of an unsupervised model since we do not know the true Y values*

UNSUPERVISED LEARNING PROBLEM

- *Even though there is no response (target) variable we still want to understand the*
  - *relationship among variables (columns)*
  - *relationship among observations (rows)*

UNSUPERVISED LEARNING METHODS

- *Clustering*

  Find groups of observations with common characteristics (values or categories)

UNSUPERVISED LEARNING METHODS

- *Clustering*

  Find groups of observations with common characteristics (values or categories)

- *Principal Component Analysis (PCA)*

  Identify new variables before clustering, supervised learning modeling, dimensionality reduction, data visualization

Classification problem - Example

- *Response will pay, will not pay*
- *Borrowers attributes*
  - *age group*
  - *gender*
  - *location*
  - *ses*
  - *student*
  - *married*

Clustering problem - Example

- *Group*
- *Borrowers attributes*
  - *age group*
  - *gender*
  - *location*
  - *ses*
  - *student*
  - *married*