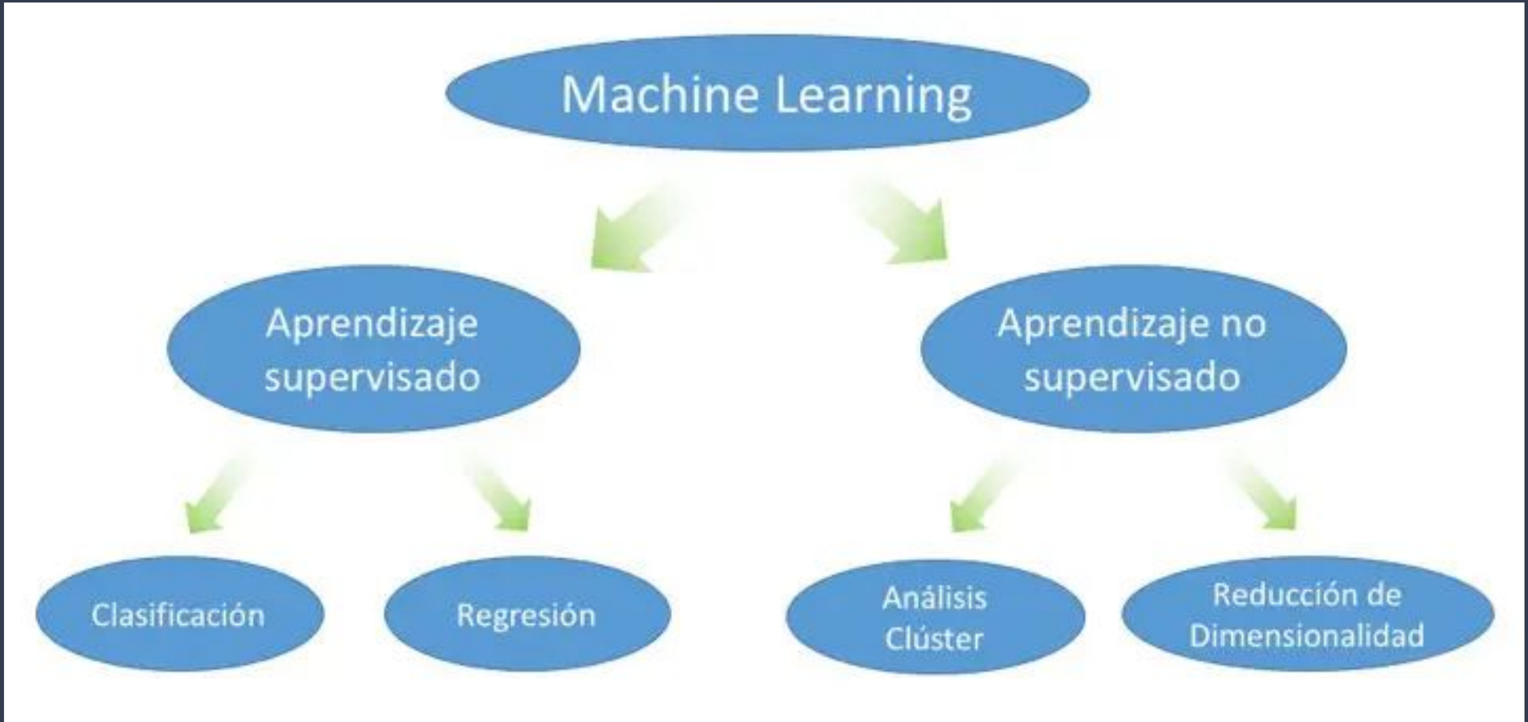
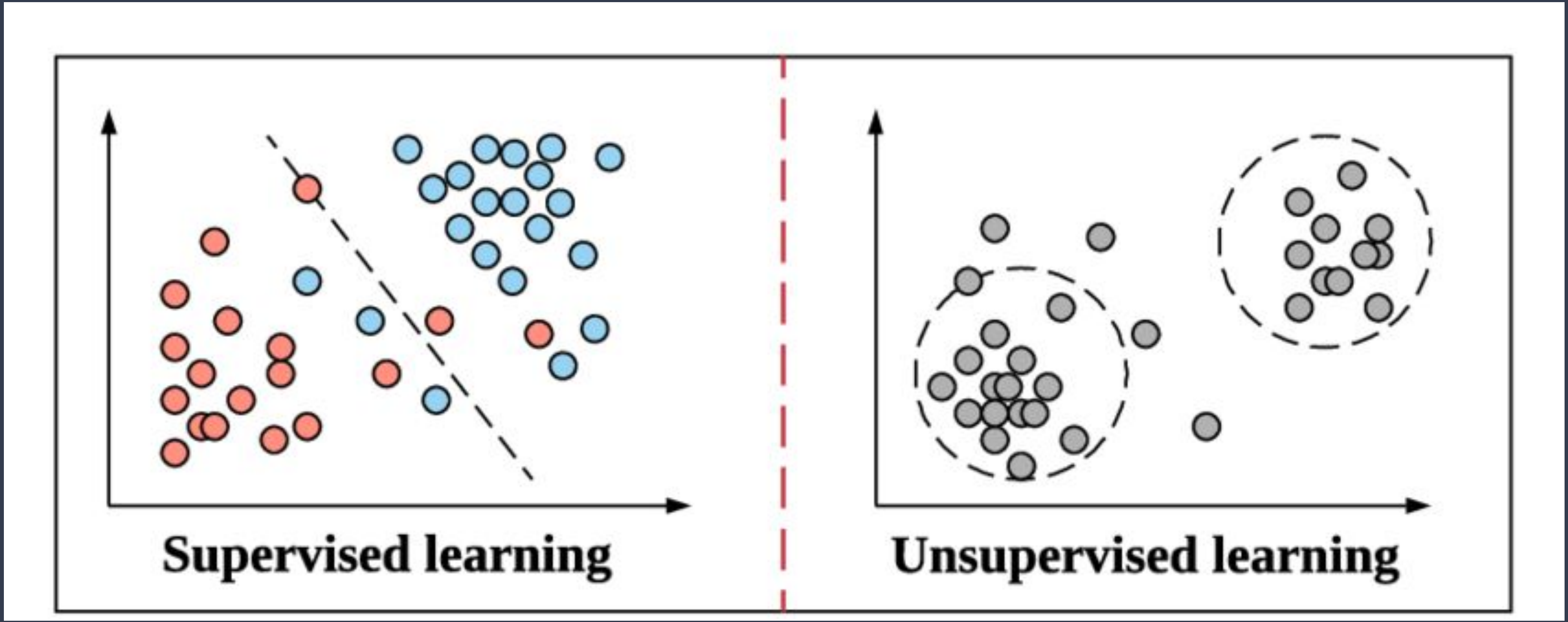
Index Class

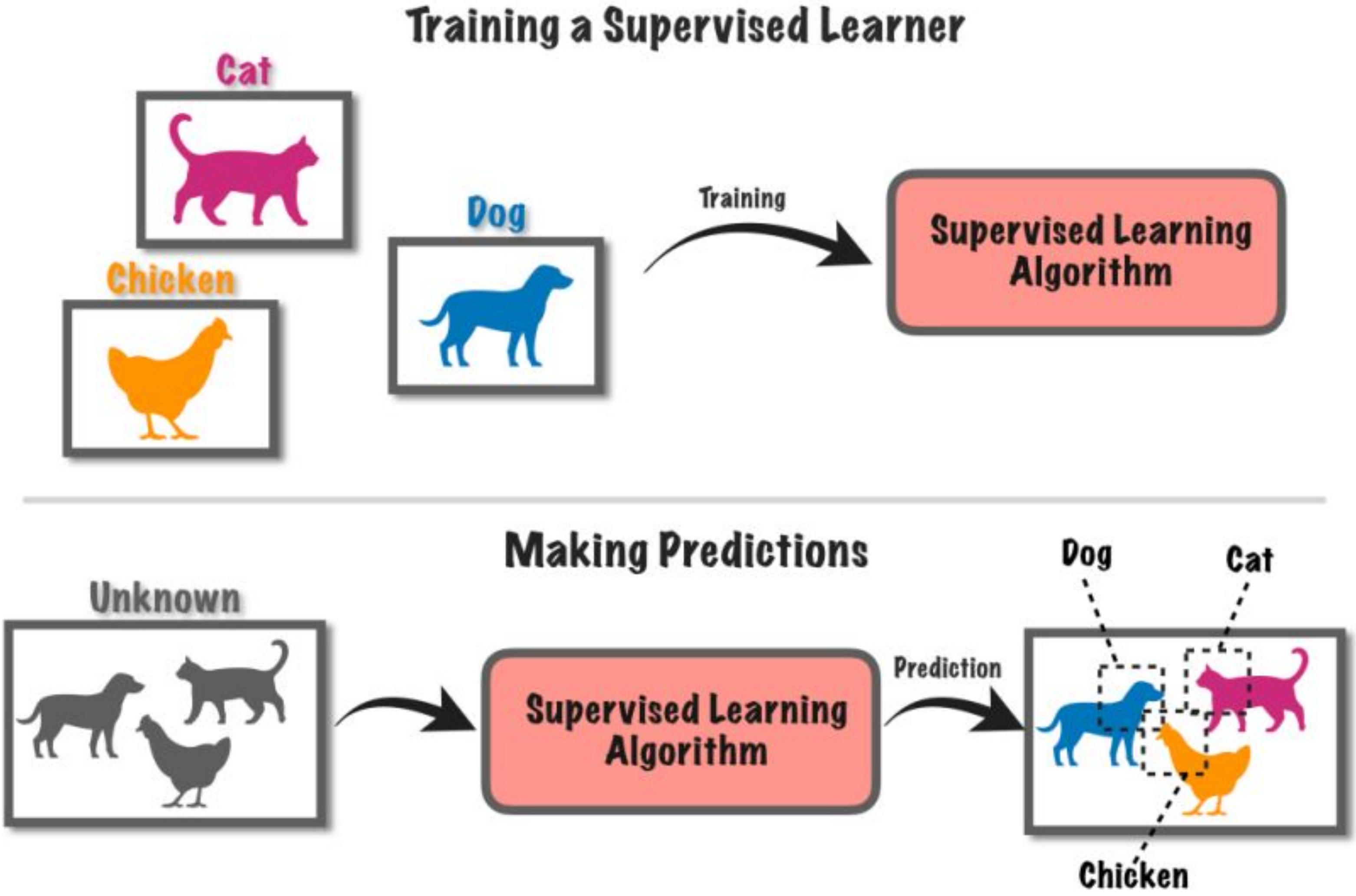
1. Statistical learning
2. Linear regression
3. Metrics
4. Estimation
5. Interpreting the estimators
6. Polynomial regression

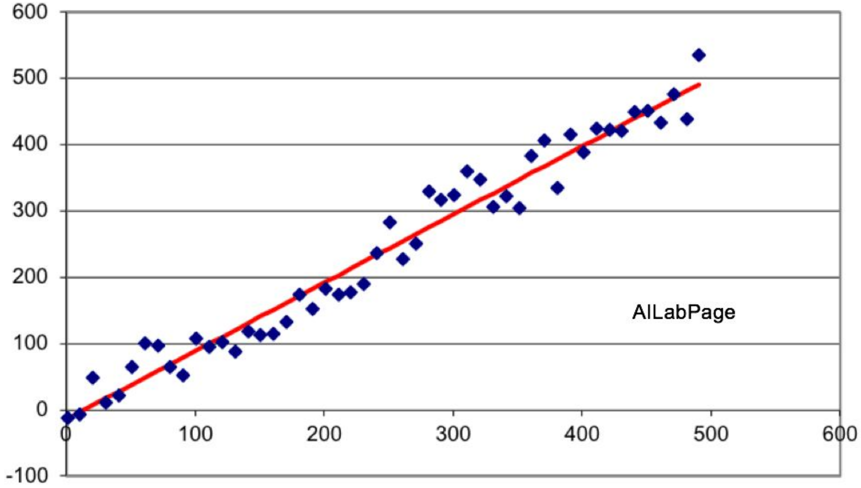
Machine learning

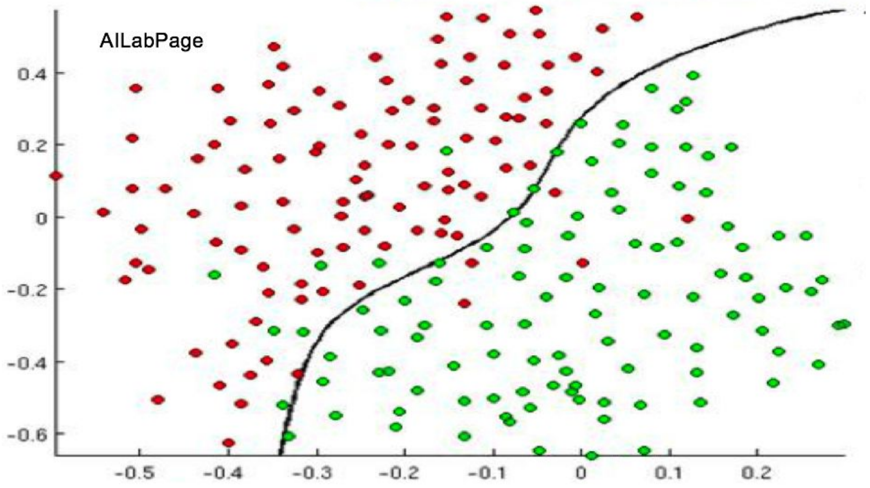


Types of algorithms




Training a Supervised Learner
Cat
Chicken
Dog
Training
Supervised Learning Algorithm
Making Predictions
Unknown
Supervised Learning Algorithm
Prediction
Dog
Cat
Chicken





Regression

The system attempts to predict a value for an input based on past data.

Example – 1. Temperature for tomorrow

Classification

In classification, predictions are made by classifying them into different categories.

Example – 1. Type of cancer 2. Cancer Y/N

Machine learning

Machine learning means learning from data:

- We have a quantitative outcome (regression) or categorical outcome (classification)
- We want to predict the *outcome* based on a set of *features* (supervised)
- We have a *training set*
- We build a prediction model for new unseen objects. The objective is to predict accurately

Vocabulary

Outcome: Usually denoted by Y

Features: Usually denoted by X (X is a vector of k features)

Training set: (x1, y1), ..., (xn, yn)

Objective: get a good prediction of Y called $\hat{Y} = f(X)$.

LOSS FUNCTION for penalizing errors (**cost function**)
Squared loss error $(Y - f(X))^2$ (Y- b- a*X1)

Simple Linear regression

$$\underset{\substack{\uparrow \\ \text{Dependent} \\ \text{Variable}}}{Y_i} = \overset{\substack{\text{Constant/Intercept} \\ \downarrow}}{\beta_0} + \underset{\substack{\uparrow \\ \text{Slope/Coefficient}}}{\beta_1} \overset{\substack{\text{Independent} \\ \text{Variable} \\ \downarrow}}{X_i}$$

Metrics

R2:

$$R^2 = 1 - \frac{\Sigma(y - \hat{y})^2}{\Sigma\left(y - \bar{y}\right)^2}$$

RMSE:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

Regresión lineal simple

Linear Regression

A linear regression model assumes that Y is linear in the inputs X:

$$f(X) = \beta_0 + \sum_{j=1}^{p} X_j \beta_j$$

$$Y = f(X) + \varepsilon$$

Basic assumptions on errors: $\varepsilon \sim N(0, \sigma^2)$

- Independent
- Mean zero
- Constant variance

Predicting new data

Given a new set of features (X_nuevo), we can predict the outcome as:

$$\hat{Y}_{nuevo} = \hat{\beta} X_{nuevo} = \hat{\beta}_0 + \hat{\beta}_1 X_{1,nuevo} + \cdots + \hat{\beta}_p X_{p,nuevo}$$

Features types

Quantitative - Continuous variables:

- Transformations: log, square root…
- Expansions: ^2, ^3, …
- Interactions: X3 = X1*X2

Qualitative - Categorical variables:

- Dummy coding of the levels. 1 variable with K categories -> K dummy variables

Interpreting estimators

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 D_1 + \hat{\beta}_3 D_2$$

X1 is a continuous variable:
- sign
- size
- marginal effect

D1, D2 are the dummys of a categorical variable with 3 levels:
- reference category D3

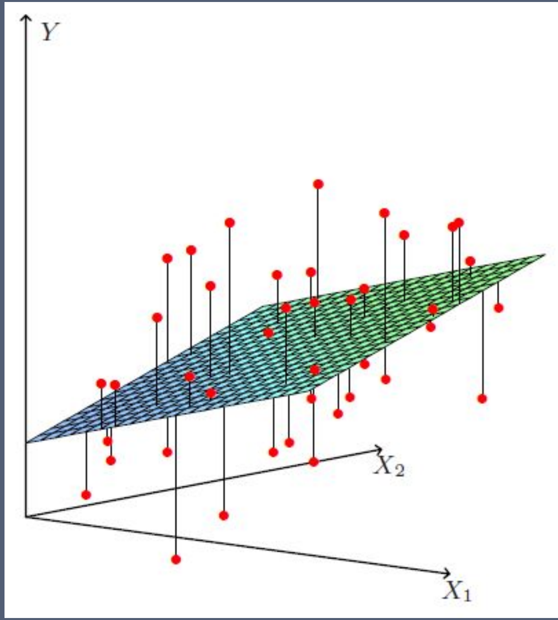
Estimation: Least Squares in LR

$$\mathrm{RSS}(\beta) = \sum_{i=1}^{N}(y_i - x_i^T\beta)^2.$$

$$\mathrm{RSS}(\beta) = (\mathbf{y} - \mathbf{X}\beta)^T(\mathbf{y} - \mathbf{X}\beta)$$

$$\frac{\partial \mathrm{RSS}}{\partial \beta} = -2\mathbf{X}^T(\mathbf{y} - \mathbf{X}\beta)$$

$$\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$



Polynomial regression

f(x) is a polynomial of order k:

$$f(X) = \sum_{j=0}^{k} \beta_j X^j$$