

# Supervised Learning

Diego ROJAS

May 28, 2022

## 1 Introduction

Supervised learning is a method of machine learning which uses the *inputs* to predict the *outputs*.

**Definition:** (**Predictor**): Synonym for input.

**Definition:** (**Responses | Dependent variables**): Synonym for output. Can be a *quantitative (discrete)* or *qualitative (categorical)* measurement. There also exists *ordered categorical* variables which refer to outputs such as "small", "large", "medium".

**Definition:** (**Regression**): When we predict quantitative outputs.

**Definition:** (**Classification**): When we predict qualitative outputs.

**Definition:** (**Training data**): A dataset  $(x_i, y_i)$  or  $(x_i, g_i), i = 1, \dots, N$  with which we construct our prediction rules.

**Definition:** (**Binary coded target**): A mean of representing a 2-valued categorical output  $G$  by encoding it as  $Y$  and then treat it as a quantitative output.

## 2 Linear Model

**Definition:** (Linear Model): Given a vector of inputs  $X^\top = (X_1, \dots, X_p)$  we predict the output  $Y$  via the model

$$\hat{Y} = \hat{\beta}_0 + \sum_{j=1}^p X_j \hat{\beta}_j \quad (1)$$

Where  $\hat{\beta}_0$  is the *intercept* or *bias*. By including 1 in  $X$  and  $\hat{\beta}_0$  in the vector of coefficients  $\hat{\beta}$  we can write

$$\hat{Y} = X^\top \hat{\beta} \quad (2)$$

## 3 Least Squares