Goal: To predict the value of one or more continuous target variables t given d-dim x input vectors.

LINEAR regression : most simple approach

SPECIFIC EXAMPLE OF Linear regression:

Type of variables : independent $x \in \mathbb{R}^d$

dependent $y \in \mathbb{R}$

We assume a linear relationship: $y = x^T\beta$

$$y = \sum_{i=1}^{d} x_i \beta_i$$

→ FXN: polynomial curve fitting.

* NOTE THIS APPROACH IS A NON-BAYESIAN/Frequentist.

Applying the frequentist notion of probability to the random values of the observed variables t_n.

vector form

~~EXTRA~~ STATISTICAL MODEL

Generalized setting: $y = \beta_0 + \sum_{j=1}^{d} \beta_j x_j$, where β_0 = intercept, bias term

Now introduce homogenous coordinates: to rewrite generalized equation in a compact form. This way you don't worry about the ~~offset~~ offset from the origin on the y-axis $x_0 = 1$

$$\Rightarrow y = x^T\beta \quad x, \beta \in \mathbb{R} \rightarrow d+1 \text{ dimensions}$$

Now with noisy observations in data: $(x_i, y_i) \sim p(x, y \mid \beta)$ are distributed according to some β.

You can assume the equation with noise that is iid. in which case, you are also trying to model the noise by predicting β.

$y_i = x_i^T\beta + \varepsilon_i$, where ε_i is noise with iid.

ε_i : $1 \leq i \leq n$ is iid.

So the question changes to; what is the probability of the ~~observed~~ observing a noise ε_i given a β.

$$P(\varepsilon_i \mid \beta) = P(\underbrace{y_i - x_i^T\beta}_{\text{residuals}} \mid \beta)$$

If you know what ε_i, then you would know how $y_i - x_i^T\beta$ is distributed.

y_i = True value

$\hat{y}_i = x_i^T\beta_i$: value of dependent variable/(prediction) predicted by the model.

Estimation problem involves determing β from observations (training data),

$$Z = \{(x_i, y_i),\ 1 \leq i \leq n\}$$