$f : x \mapsto t$

True underlying relationship is :

$$t = \sin(2\pi x)$$

But, data is noisy.

$$t_n = \sin(2\pi x) + \varepsilon_n$$

where

$$\varepsilon_n \sim \mathcal{N}(0, \sigma^2)$$

Each point (x_n, t_n) represents one observation from the joint distribution $p(x, t)$

$$p(t \mid x) = \mathcal{N}(t \mid \sin(2\pi x), \sigma^2)$$

<u>The Big Idea</u>

We want to find a function $[y(x, w)]$ that approximates the true relationship b/w input x and output t.

$$y(x, w) = w_0 + w_1 x^1 + w_2 x^2 + \cdots + w_M x^M = \sum_{j=0}^{M} w_j x^j$$

Why it's called "Linear Model"

Even though $y(x, w)$ is nonlinear in x, it's linear in the parameter w.

$$y(x, w) = \underbrace{[1, x, x^2, \dots, x^M]}_{\phi(x)^T} \underbrace{\begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_M \end{bmatrix}}_{w}$$

$$= \phi(x)^T w$$

$$\Phi = \begin{bmatrix} 1 & x_1 & x_1^2 & \dots & x_1^M \\ 1 & x_2 & x_2^2 & \dots & x_2^M \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n & x_N^2 & \dots & x_N^M \end{bmatrix}$$

→ Each row is different sample from the data.

→ Each column is a diff. version of the same input.

Error function

Sum of Squares Error :

$$E(w) = \frac{1}{2} \sum_{n=1}^{N} \{ y(x_n, w) - t_n \}^2$$

WRt, $y = \phi w$,
Now let $t := [t_1, t_2, \ldots, t_n]$

$$E(w) = \frac{1}{2} (\phi w - t)^T (\phi w - t)$$

$$\nabla_w E(w) = \phi^T (\phi w - t) = 0$$

DO WHY, HOW

$$\phi^T \phi w = \phi^T t$$

$$w^* = (\phi^T \phi)^{-1} \phi^T t$$