③ Polynomial Regression

p- power ; d- dimension. The function $f_w(x)$ has $\binom{d+p}{p}$ terms

For example, if $p=2$, $d=2$.

$$f_w(x) = w_0 + w_1 x_1 + w_2 x_2 + w_3 x_1 x_2 + w_4 x_1^2 + w_5 x_2^2$$

matrix P is $m \times \binom{d+p}{p}$; $\vec{y} \in \mathbb{R}^m$

$$w^* = P^T (PP^T)^{-1} \vec{y} \text{ , where } P = \begin{bmatrix} 1 & x_{11} & x_{12} & x_{11}x_{12} & x_{11}^2 & x \\ 1 & x_{m1} & x_{m2} & x_{m1}x_{m2} & x_{m1}^2 & x \end{bmatrix}$$

④ Ridge Regression. — a kind of Linear Regression.

but $d \gg m$. which means numerous variables with only a few samp

$\Rightarrow X^T X$ may be not invertible.

Let $J(\bar{w}) = \sum_{i=1}^{m} (y_i - \bar{w}^T \bar{x}_i)^2 + \lambda \sum_{j=0}^{d} w_j^2$

$$= (X\bar{w} - \vec{y})^T (X\bar{w} - \vec{y}) + \lambda \bar{w}^T \bar{w}$$

$$= \bar{w}^T (X^T X + \lambda I) \bar{w} - 2\bar{w}^T (X^T y) + y^T y$$

Differentiate it and let $\nabla_{\bar{w}} J(\bar{w}) = 0$.

$$\Rightarrow \bar{w}^* = (X^T X + \lambda I_{d+1})^{-1} X^T \vec{y}$$

$$= X^T \underbrace{(XX^T + \lambda I_m)}_{\text{why invertible}}{}^{-1} \vec{y}$$

why equation is true.

A few useful propositions:

① If a matrix $\in \mathbb{R}^{n \times n}$ is positive-definite / negative-definite, then it's invertible.

x is a vector $x^T A x > 0$ / $x^T A x < 0$.

② Woodbury formula : $(I + UV)^{-1} = I - U(I + VU)^{-1} V$