How to optimize linear regression?

We want to minimize the loss function, commonly we use the square error

$$E(\vec{w}) = \sum (y_i - (\vec{x}_i^T \vec{w}))^2$$
$$= \| \vec{y} - X\vec{w} \|^2$$
$$= (\vec{y} - X\vec{w})^T (\vec{y} - X\vec{w})$$
$$= \vec{w}^T X^T X \vec{w} - 2\vec{y}^T X \vec{w} + \vec{y}^T \vec{y}$$

Note: the above function is equadratic.

We can use the gradient $= 0$ to solve for the optimal values.

$$\nabla E = \left[\frac{dE}{dw_0} \quad \frac{dE}{dw_1} \quad \cdots \quad \frac{dE}{dw_n} \right]$$

The equate each value to zero

$$\frac{dE}{dw_0} = 0 \quad \frac{dE}{dw_1} = 0 \quad \cdots \quad \frac{dE}{dw_n} = 0$$

Quadratic Function:

$f(x) = \vec{x}^T A \vec{x} + \vec{b}^T \vec{x} + c$, where $\vec{x}^T A \vec{x}$ is positive definite.

f will always open up. When we get the Hessian matrix of 2nd order partial derivatives

The Hessian Matrix $= A$.

Continuing the gradient calculations ...

$$\nabla E = 2(X^T X)\vec{w} - 2X^T\vec{y} = 0$$

$$\Rightarrow \vec{w}^* = (X^T X)^{-1} X^T \vec{y}$$

$$= P\vec{y}$$

This $P = (X^T X)^{-1} X^T$ is called the pseudo-inverse.

1D Gaussian Distribution.

The pdf with params μ, σ^2 is

$$f(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)$$

Where $\mu = E(x) = \int x f(x)\,dx$

and $\sigma^2 = E((x-\mu)^2) = \int (x-\mu)^2 f(x)\,dx$