HW4

Wednesday, October 14, 2020 8:01 PM

1. $Loss(w) = MSE(w) + \alpha w^T w = \frac{1}{N}(Y - wX)(Y - wX)^T + \alpha w^T w$

$= \frac{1}{N}[YY^T - 2YX^Tw^T + wxx^Tw^T] + \alpha w^T w$

$= \frac{1}{N}[YY^T - 2YX^Tw^T + wx^Txw^T + 2\alpha w^T w]$

$= \frac{1}{N}[w(x + \sqrt{\alpha}I)^T(x + \sqrt{\alpha}I)w^T - 2x^Tw^TY + YY^T]$

$\nabla Loss(w) = \nabla MSE(w) + 2\lambda w$

$= 2x^Txw - 2x^TY + 2\lambda w = 0$

$\Rightarrow x^Txw + \lambda w = x^TY$

$(x^Tx + \lambda I)w = x^TY$

$\Rightarrow \quad w = (x^Tx + \lambda I)^{-1}x^TY$

2. $MSE(w) = \frac{1}{N}[Y - relu(wx)]^T[Y - relu(wx)]$

$= \frac{1}{N}\left[Y^TY - Y^T relu(wx) - (relu(wx))^TY + [relu(wx)]^T[relu(wx)]\right]$

Because $relu(wx)^T relu(wx) \neq relu[(wx)^T(wx)]$,

This is not a quadratic form.

$\nabla MSE(w) = relu(x^T) \cdot relu(wx) + relu(x^Tw^T) \cdot relu(x) - 2relu(Y^Tx)$

$w \leftarrow w - \alpha \nabla MSE(w)$

$= w - 2relu(x^T)relu(wx) + 2relu(Y^Tx)$

3. Give up . . . .