

what

b. L^2 正则化.

$$\bar{J}(w; X, y) = \frac{\alpha}{2} w^T w + J(w; X, y)$$

$$\nabla \bar{J}(w; X, y) = \alpha w + \nabla J_w(w; X, y)$$

$$w \leftarrow (1 - \frac{1}{2}\alpha)w - \frac{1}{2}\nabla_w J(w; X, y)$$

加入权重衰减会引起学习规则的修改, 即在每步执行通常的梯度更新之前先收缩 w .

7.1

where

泰勒近似 $\bar{J}(w) = J(w^*) + \frac{1}{2}(w - w^*)^T H (w - w^*)$

$$\nabla_w \bar{J}(w) = H(w - w^*)$$

设最小值在 \tilde{w} 上取得并加上正则项.

$$\alpha \tilde{w} + H(\tilde{w} - w^*) = 0$$

$$\tilde{w} = (H + \alpha I)^{-1} H w^*$$

因 $H = Q \Lambda Q^T$ 有

$$\tilde{w} = Q(\Lambda + \alpha I)^{-1} \Lambda Q^T w^*$$

可以看到权重衰减由 H 的特征向量所定义的轴收缩 w^* , $\frac{\lambda_i}{\lambda_i + \alpha}$.

~~若~~

沿 H 特征值较大方向 ($\lambda_i \gg \alpha$) 正则化影响很小.

($\lambda_i \ll \alpha$) 分量将会收缩到几乎为 0.









