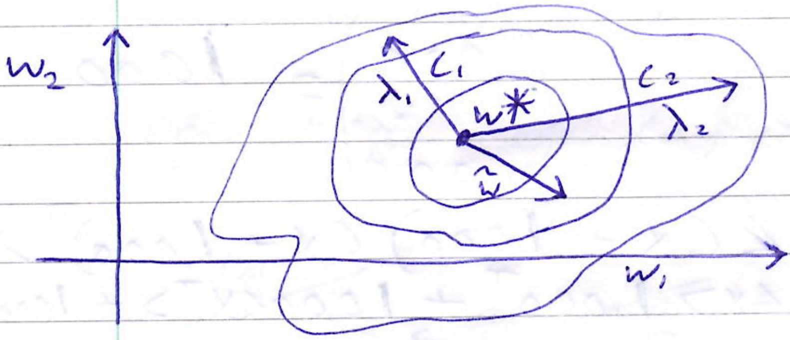
Machine Learning

09/10/17

$$w(t+1) = w(t) - y \nabla E(w(t))$$

Learning rate   gradient



Non-linear case

$$w = w^* + \tilde{w}$$

Error = E

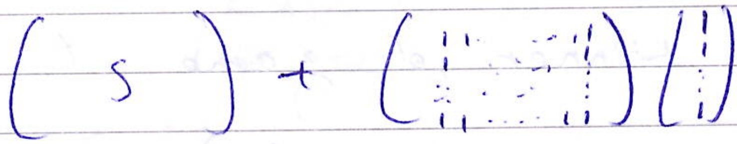
$$E(w) = E(w^* + \tilde{w}) =$$
$$= E(w^*) + \nabla E(w^*) \cdot (\tilde{w}) + \tfrac{1}{2} \tilde{w}^T \underbrace{\nabla^2 E(w^*)}_{\text{Heissen}} \tilde{w} + 0$$

$$DE \cdot w^2 = \sum_i \frac{\partial E}{\partial w_i} \tilde{w}_i$$

$$\frac{1}{2} w^T (\nabla^2 E) \tilde{w} = \frac{1}{2} \sum_{ij} \left(\frac{\partial^2 E}{\partial w_i \partial w_j} \right) \tilde{w}_i \tilde{w}_j$$

if you have multiple variables that you need to adjust [n-dim] use:

$$w(t+1) = w(t) - y^P \nabla E(w(t))$$



$$\hat{z} = w \cdot x \qquad E = \langle \tfrac{1}{2} |\hat{z} - z|^2 \rangle$$

$$\nabla E = \langle (\hat{z} - z) x \rangle$$

$$\nabla^2 E = \langle x x^T \rangle = H = C_{xx}$$