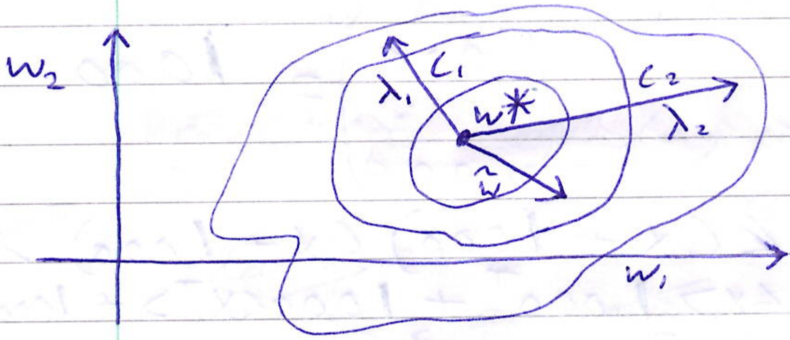
Machine Learning

09/10/17

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

Learning rate   gradient



Non-linear case

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

Error = E $E(w) = E(w^* + \tilde{w}) =$

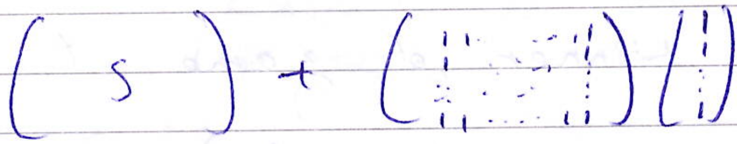
$$= E(w^*) + \nabla E(w^*) \cdot (\tilde{w}) + \tfrac{1}{2} \tilde{w}^T \underbrace{\nabla^2 E(w^*)}_{\text{Heissen}} \tilde{w} + O$$

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

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

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

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



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

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

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

$$h_{ij} = \frac{\partial^2 E}{\partial w_i \partial w_j} = \langle x_i x_j \rangle = \frac{1}{m} \sum_p x_i^{(p)} x_j^{(p)}$$

$$\hat{X} = X + \begin{pmatrix} 1000 \\ \cdots \\ \cdots \\ 1000 \end{pmatrix} \qquad \begin{pmatrix} 1000 \\ \cdots \\ \cdots \\ 1000 \end{pmatrix} = \mathbf{1000}$$

$$\tilde{H} = \langle \tilde{X} \tilde{X}^T \rangle = \langle (X + \mathbf{1000})(X + \mathbf{1000})^T \rangle$$
$$= \langle X X^T \rangle + \cancel{\langle X \rangle \mathbf{1000}^T} + \cancel{\mathbf{1000} \langle X^T \rangle} + 1000$$

assume $\langle X \rangle = 0$
as there is zero mean.

Make (S) input zero mean and make input unit varience

$$h_{ij} = \frac{\partial^2 E}{\partial w_i^2} = \langle x_i^2 \rangle = X^{-2} + VAR(x_i)$$

if we have n "experts"

~~$w_i(t)$ = weight for i for t^{th} game (± 1)~~

$w_i(t)$ = weight for expert (i) at time (t)

$x_i(t)$ = prediction for i for t^{th} game ($\pm$ 1)

$$P(t) = \text{sign}\left(\sum_i w_i(t)\, x_i(t)\right)$$

$z(t) = \pm 1$ winner of game t

$$v_i(t) = \frac{\#\ \text{correct}\ (0 \ldots t-1)}{t}$$

$v_i(t+1) = \ldots \quad v_i(t)$ [Continued next lecture]

S = matrix of small numbers.