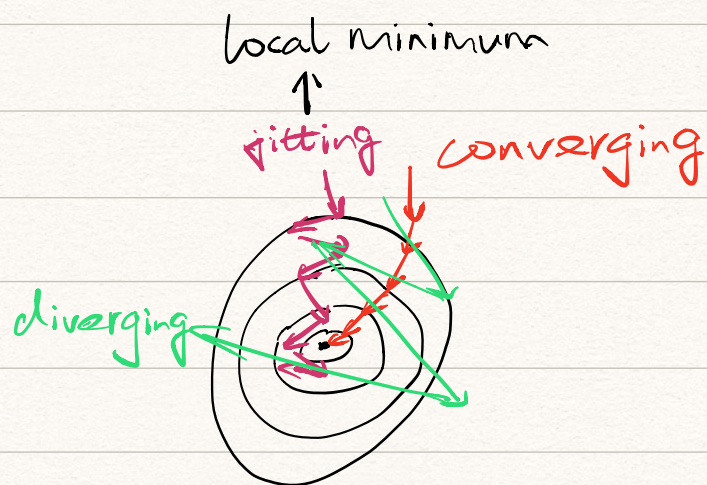


Convex Optimization

convex loss function

Streetlight effect: caveat: Neural Network error surface is generally not convex



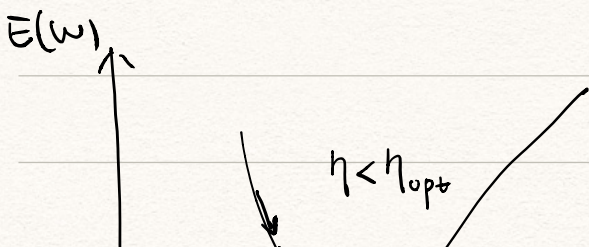
convergence rate:
$$R = \frac{|f(x^{(k+1)}) - f(x^*)|}{|f(x^{(k)}) - f(x^*)|}$$

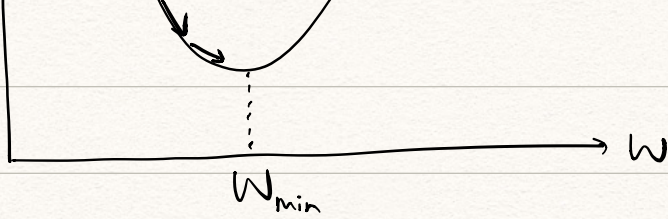
(x^* is the optimal value of x)

二次的
Convergence for quadratic surfaces

Minimize $E = \frac{1}{2}aw^2 + bw + c$

$$w^{(k+1)} = w^{(k)} - \eta \frac{dE(w^{(k)})}{dw}$$





$$E = E(w^{(k)}) + E'(w^{(k)})(w - w^{(k)}) + \frac{1}{2}E''(w^{(k)})(w - w^{(k)})^2$$

$$w_{\min} = w^{(k)} - E''(w^{(k)})^{-1} E'(w^{(k)})$$

$$\frac{dE(w^{(k)})}{dw} = E'(w^{(k)})$$

$$\begin{aligned} w^{(k+1)} &\leftarrow w^{(k)} - \eta \nabla_w E \\ w_i^{(k+1)} &= w_i^{(k)} - \eta \frac{dE(w_i^{(k)})}{dw} \end{aligned}$$