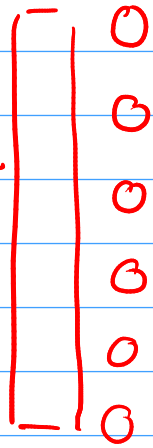


# Some few other techniques to combat overfitting

add gaussian noise to feature vector to predict stuff from ①



→ target same

② OR add gaussian noise [hidden layers] here instead (2014/15 paper)

output of

[mean of the datapoint] target can be considered the same

$$\omega^{(k+1)} = \omega^{(k)} - \eta (\nabla E(\omega) |_{\omega=\omega^{(k)}}) \quad [① \& ②]$$

be considered the same

⇒ more datapoints!

$$\omega^{(k+2)} = \omega^{(k+1)} - \begin{bmatrix} n \\ 0 \\ s_e \end{bmatrix} - \eta (\nabla E(\omega) |_{\omega=\omega^{(k)}}) \quad [③]$$

(3) add noise to the weights themselves

① Seems nice

③ Sounds like WTF but works

② Fine

## Another technique

$$(x^{(1)}, t_1) \quad (x^{(2)}, t_2) \quad \dots \quad (x^{(m)}, t_m)$$

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 & \dots & 0 \end{pmatrix}$$

Crossentropy  $p^{(0)} p^1 p^3 p^4 p^5 p^6$

Estimated probabilities functions of  $\omega$  & other params

We want to maximise probability  
 $\Rightarrow$  we minimise error

but then if 010000 is wrong (scope for error in target attribute)  
we can't acc for it

if we acc. for the probability of error to be  $\epsilon$

also hyperparameter  $\left( p_1^{\epsilon/5} p_2^{1-\epsilon} p_3^{\epsilon/5} p_4^{\epsilon/5} p_5^{\epsilon/5} p_6^{\epsilon/5} \right) \rightarrow$  Softens the error function

This is our new crossentropy

from this we get an error