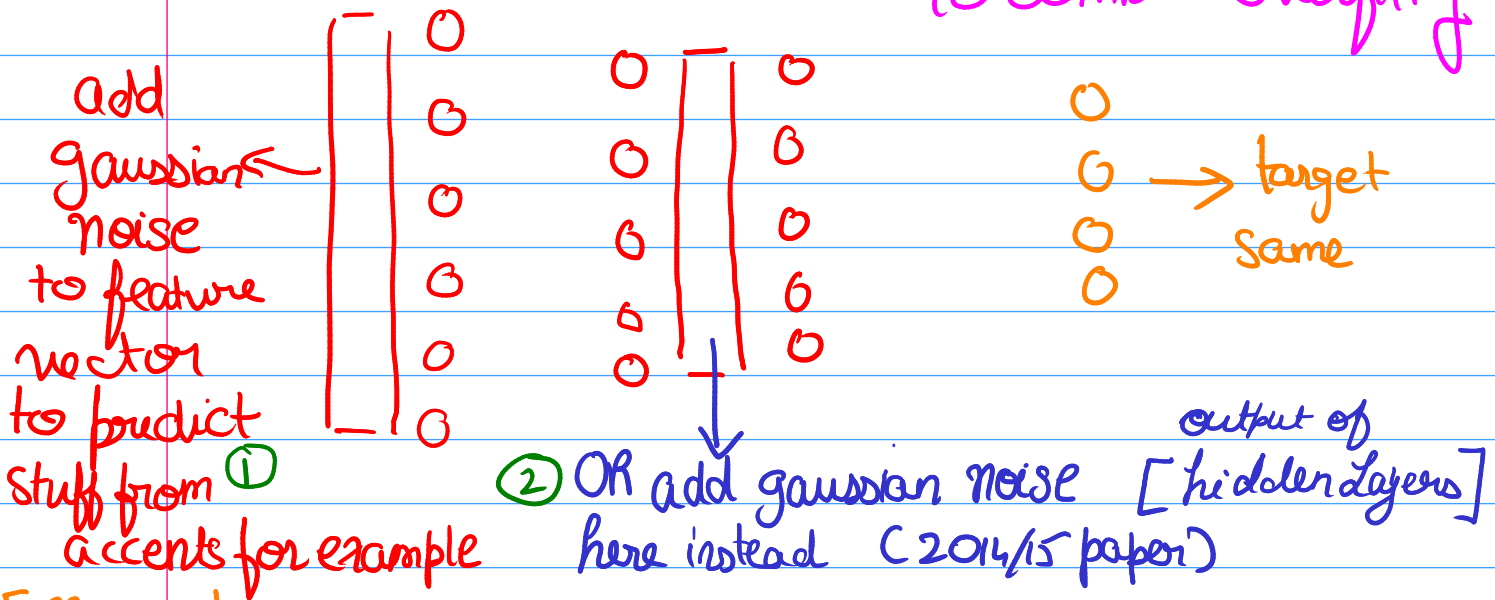
Some few other techniques to combat overfitting



[mean at the datapoint] $\omega^{(k+1)} = \omega^{(k)} - \eta \left(\nabla E(\omega) \big|_{\omega=\omega^{(k)}} \right)$ [①&②]

target can be considered the same

⇒ more datapoints!

$\omega^{(k+2)} = \omega^{(k+1)} - [\text{noise}] - \eta \left(\nabla E(\omega) \big|_{\omega=\omega^{(k)}} \right)$ ③

(3) add noise to the weights themselves

① Seems nice ③ Sounds like WTF but works
② Fine

Another technique

$(x^{(1)}, t_1) \; (x^{(2)}, t_2) \ldots (x^{(m)}, t_m)$

$(0 \quad 1 \quad 0\,0\,0 \cdots \quad 0)$

Crossentropy $p^{(0)} \; p^1 \; p_3^0 \; p_4^0 \; p_5^0 \; p_6^0$

estimated probabilities functions of ω & other params

we want to maximise probability
⇒ we minimise error

but then if 0 1 0 0 0 0 is wrong (scope for error in target attribute)
we can't acc for it

if we acc for the probability of error to be ϵ

$p_1^{\epsilon/5}$ $p_2^{1-\epsilon}$ $p_3^{\epsilon/5}$ $p_4^{\epsilon/5}$ $p_5^{\epsilon/5}$ $p_6^{\epsilon/5}$ → Softens the error function

also hyperparameter

This is our new crossentropy

from this we get our error