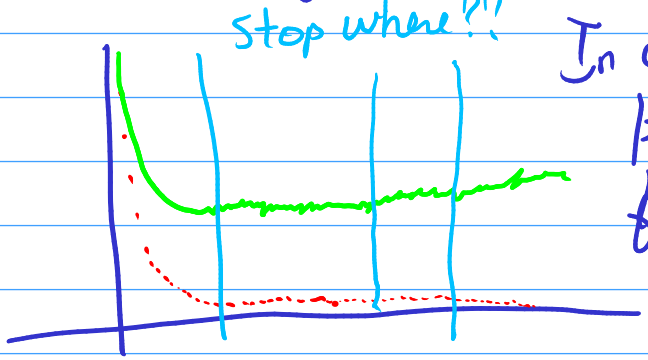
Early stopping



In a way where computational power is reduced and fit is optimal

SGD $\theta_{k+1} = \theta_k - \eta \left(\nabla E(\theta) \right)\Big|_{\theta=\theta_k}$

find training & validation error

$\theta^* : \theta_{100}$

θ_{200} has lower validation error

$\Rightarrow \theta^* : \theta_{200}$

θ_{300} has higher validation error

$\theta^* : \theta_{200}$ (Don't replace)

θ_{400} also has higher than θ_{200}

→ 100 epochs

this 100 is called "patience"

if patience = 20 ⇒ 20 iterations to observe worsening of validation error

Don't refer to this!

not necessarily increase in worsening

worsening ⇒ worse than θ^*

Sliding window over epochs

keep a counter everytime next if error ↓ otherwise ↑ 1

if counter == patience

RIP

Let n be the number of steps between evaluations.
Let p be the "patience," the number of times to observe worsening validation set error before giving up.
Let $\boldsymbol{\theta}_o$ be the initial parameters.
$\boldsymbol{\theta} \leftarrow \boldsymbol{\theta}_o$
$i \leftarrow 0$
$j \leftarrow 0$
$v \leftarrow \infty$
$\boldsymbol{\theta}^* \leftarrow \boldsymbol{\theta}$
$i^* \leftarrow i$
while $j < p$ **do**
  Update $\boldsymbol{\theta}$ by running the training algorithm for n steps.
  $i \leftarrow i + n$
  $v' \leftarrow \text{ValidationSetError}(\boldsymbol{\theta})$
  if $v' < v$ **then**
    $j \leftarrow 0$
    $\boldsymbol{\theta}^* \leftarrow \boldsymbol{\theta}$
    $i^* \leftarrow i$
    $v \leftarrow v'$
  else
    $j \leftarrow j + 1$
  end if
end while
Best parameters are $\boldsymbol{\theta}^*$, best number of training steps is i^*

Algorithm determines the best amount of time to train. The meta algorithm is a general strategy that works well with a variety of training algorithms and ways of quantifying error on the validation set.

Take epochs like for now

epochs

Update n times (iterations)