

→ learning rate ( $\alpha$ )  
 → optimizer (SGD) → Adam (adagrad + momentum)  
 → optimizer (SGD) → adagrad  
 → optimizer (SGD) → rms prop

→ model definition  
 ↳ no. of layers  
 ↳ choice of layers  
 ↳ activation fns

↑ Mathematical

model is very complex

Overfitting

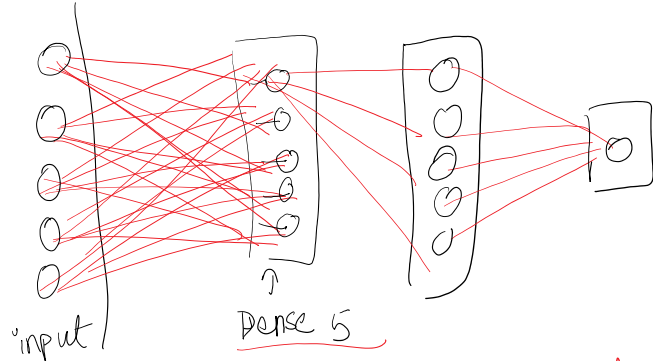
Underfitting

↑ training accuracy  
 ↳ validation/test accuracy

↳ increase layers

avoid sigmoid in hidden layers

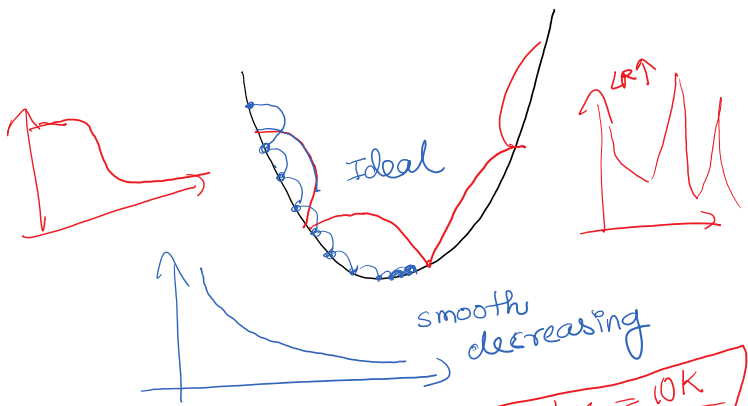
↔ Gradient Vanishing



sigmoid → classification binary

softmax → multi class classification

relu →  
 ↑  
 can use in linear reg  
 ⇓  
 relu or  
 no activation



num\_samples = 10K