Lecture 3 Word Classification and ML

Deep Learning with NLP

A neuron: Function,Parameter,Cost, Optimiser, and Gradient

— *Input: x=number of apple given by Lisa*

— *Output: y=number of banana received by Lisa*

— *Parameter: Need to be estimated*

1. Function

   — data: input,output

   — model: $Y = WX + b$

2. Parameter

   — estimate the w, b to optimise objective function

3. Cost

   — Sqare loss: $C(w, b) = \sum(y_n - \hat{y}_n)^2$

4. Optimiser

   — $argmin\ C(w, b)$

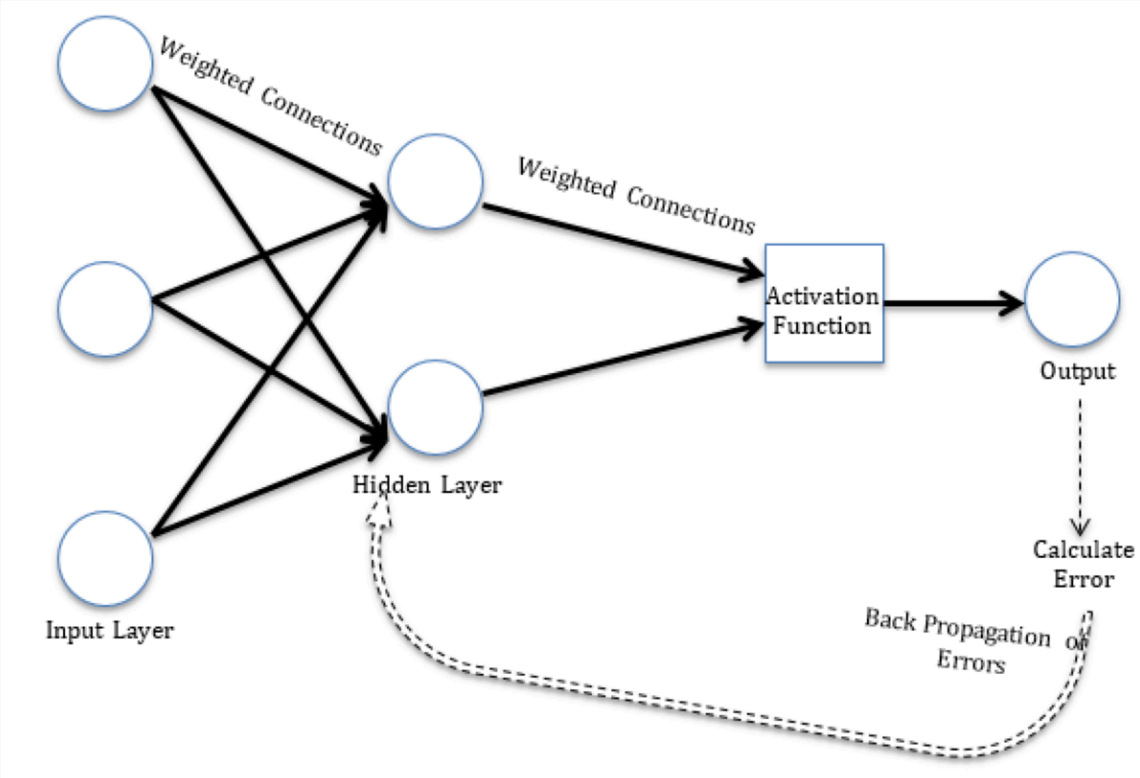
   — $w, b \in [-\infty, \infty]$

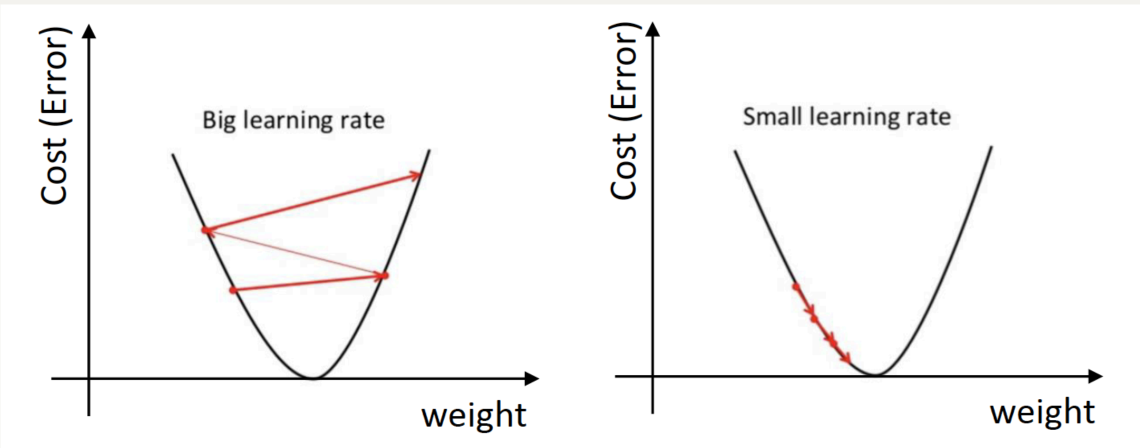
5. Gradient

$- \hat{w} = w - lr * grads$

$-$

$\hat{w}' = \hat{w} - lr * (current\ y - desired\ y) * grads(current) * existing\ input$





Parapeters vs Hyper-parameter

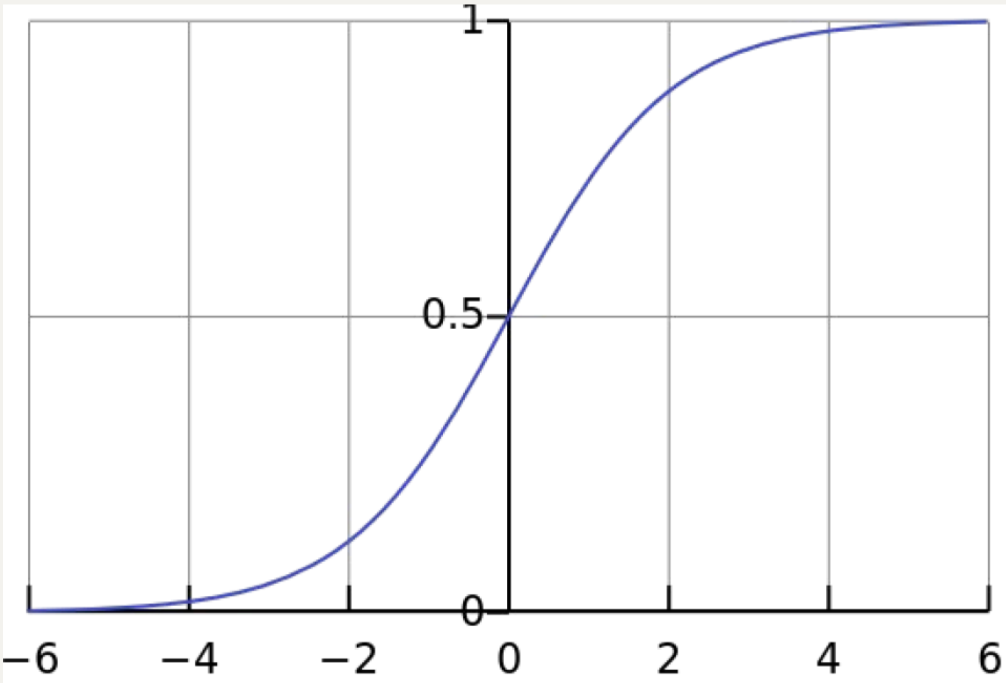
1. Parameters:
    - tunable components of model
    - learnt from training data
    - Eg. probabilities, feature weights
2. Hyper-parameters:

- Variables that controls how parameters are learnt
- Chosen a priori or tuned using held-out data
- Eg. Model size(depth,complexity), learning rate

Non-linear Neural Network

Multilayer Perceptron

1. Loss function:
    - $S(y) = O(wx + b)$
    - $O(t) = \frac{1}{1+e^{-t}}$



2. Objective function

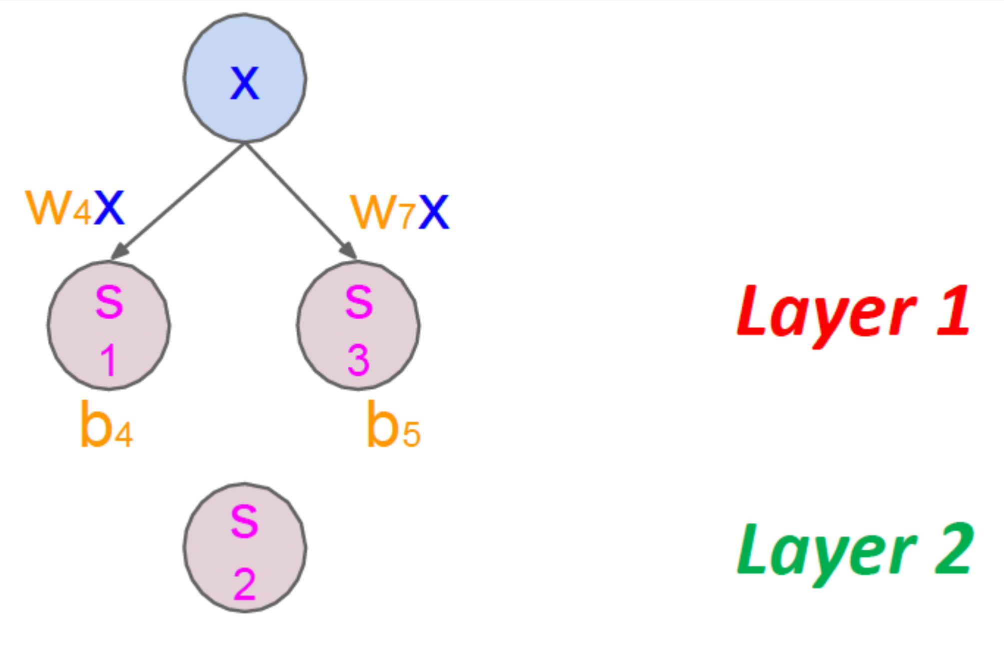
    layer 1=input features

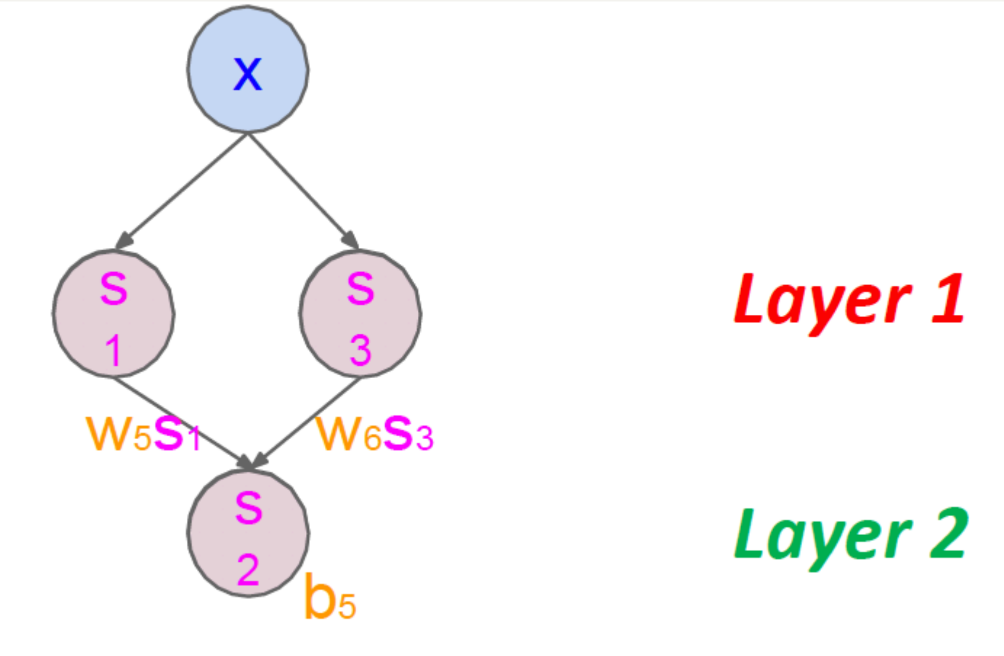
    layer 2=add and or combinations

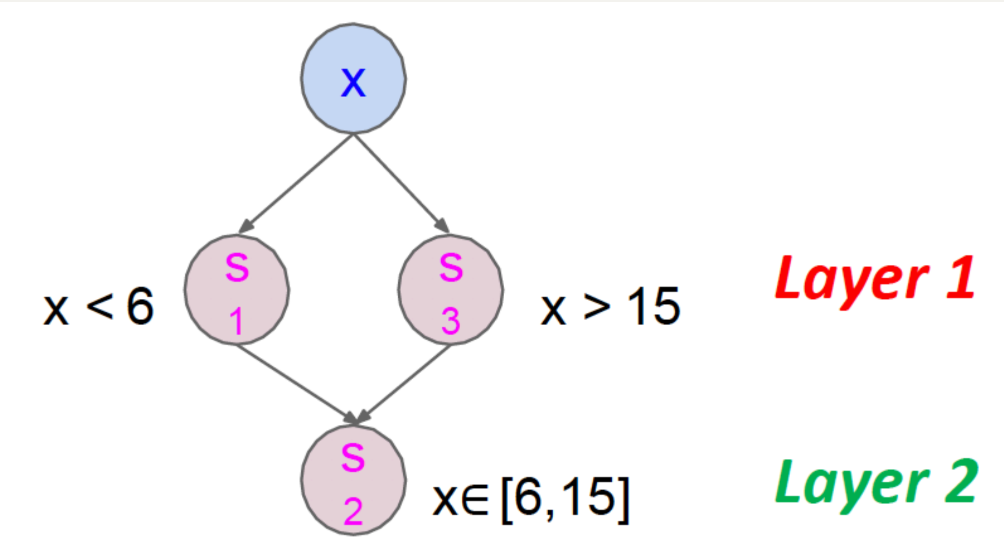
    $y = (w_1 x + b_1)S_1 + (w_2 x + b_2)S_2 + (w_3 x + b_3)S_3$

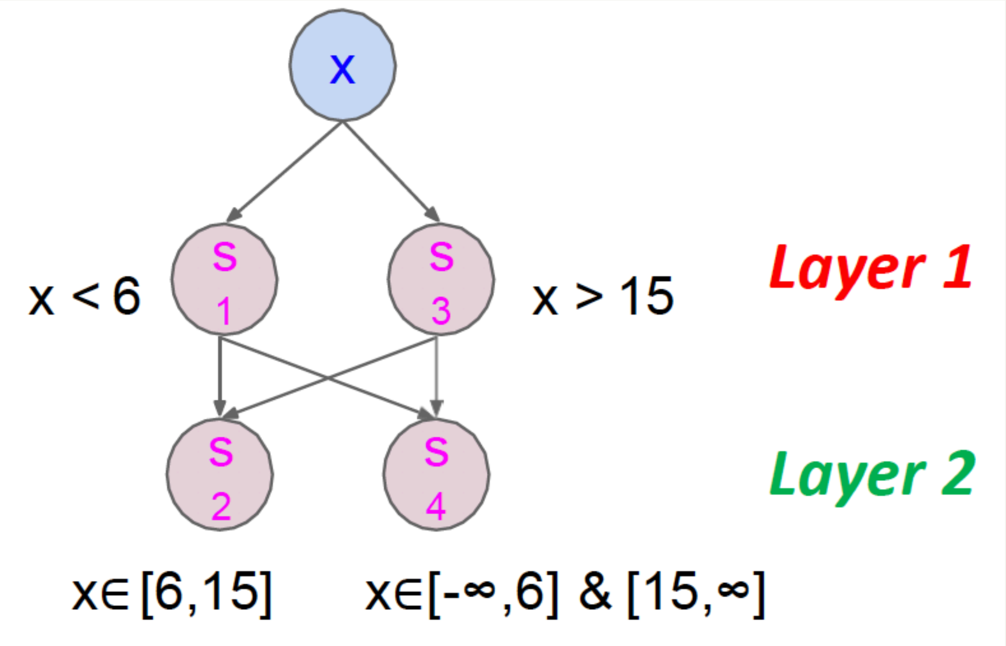
    a. Layer 1: $S_1 = O(w_4 x + b_4)$
    b. Layer 2: $S_2 = O(w_5 S_1 + w_6 S3 + b_5)$

c. Layer 1: $S_3 = O(w_7 x + b_6)$

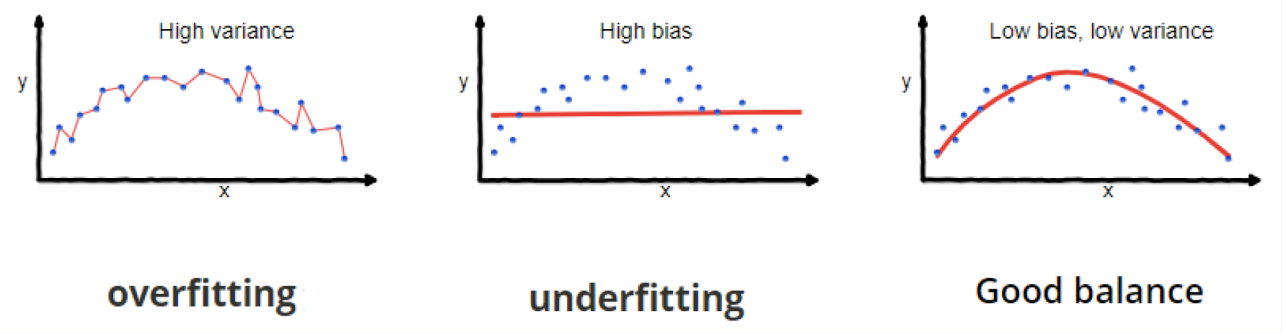


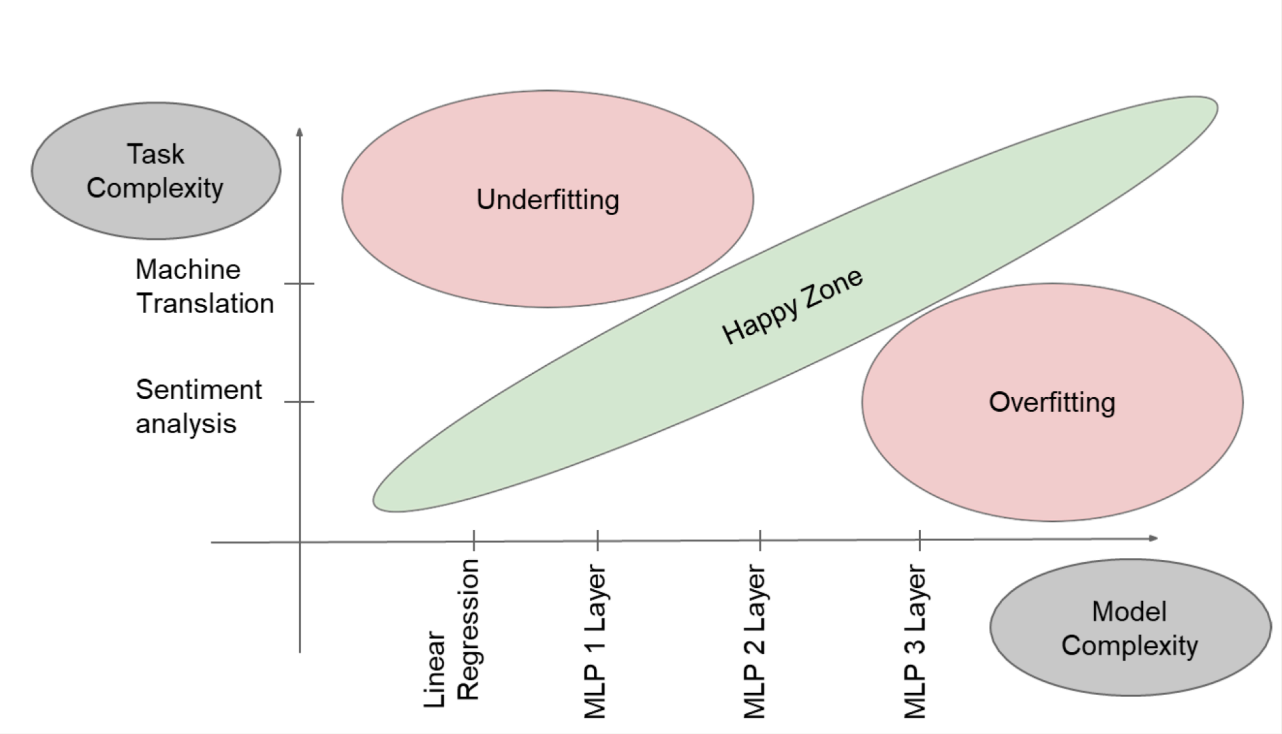





X
S 1
S 3
x < 6
x > 15
Layer 1
S 2
S 4
Layer 2
x∈[6,15]
x∈[-∞,6] & [15,∞]

Evaluation


High variance
High bias
Low bias, low variance
overfitting
underfitting
Good balance

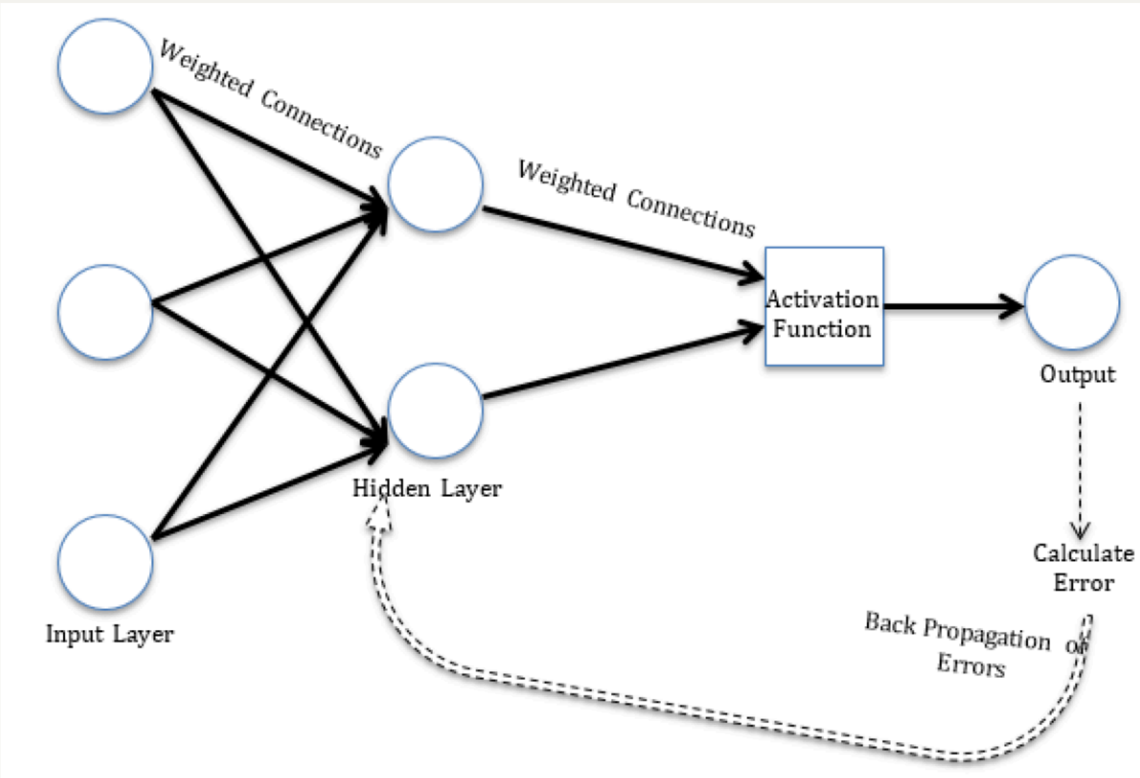

Task Complexity
Underfitting
Machine Translation
Happy Zone
Sentiment analysis
Overfitting
Linear Regression
MLP 1 Layer
MLP 2 Layer
MLP 3 Layer
Model Complexity

Score nomalization and Cost function

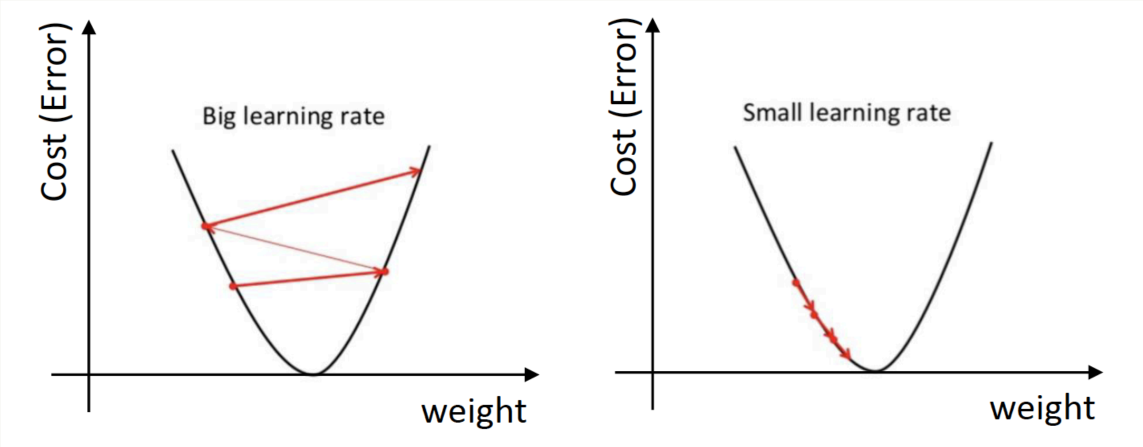
Problem type	Last-layer activation	Loss (Cost) function	Example
Binary classification	sigmoid	Binary Cross Entropy	Sentiment analysis (Positive/ Negative)
Multi-class, single-label classification	softmax	Categorical Cross Entropy	Part-of-Speech tagging Named Entity Recognition
Multi-class, multi-label classification	sigmoid	Binary Cross Entropy	Multi-topic classification, one can have multiple topics
Regression to arbitrary values	None	MSE (Mean Squared Error)	Predict house price
Regression to values between 0 and 1	sigmoid	MSE or Binary Cross Entropy	Engine health assessment where 0 is broken, 1 is new

Parameter update

1. Backpropagation



2. Gradient(=slope)



3. Gradient decent optimisation

Summary