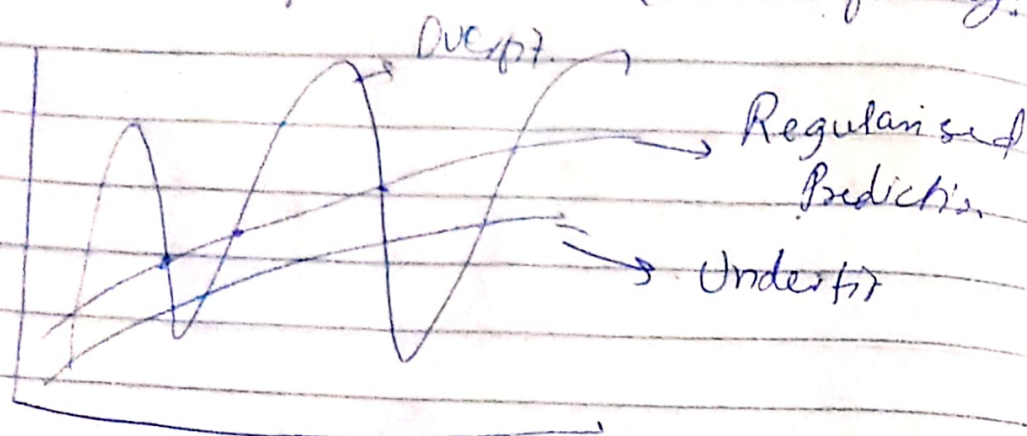


Day 6

## Regularization

- It helps us to control our model capacity
- Become able to generalize on test data
- It helps to avoid overfitting.



- Our goal is to obtain the Regularised Predictions
- Updating our loss and weight to include regularization:

$$L_i = -\log\left(\frac{e^{s_{ij}}}{\sum_j e^{s_{ij}}}\right)$$

The our entire training set :

$$\lambda = \frac{1}{N} \sum_i L_i \quad R(W) = \sum_i \sum_j W_{ij}^2$$

Most common regularization →

$L_1$ ,  $L_2$ .

$L_2$  → discourages large weights in our  $(W)$

→ It avoids dependence of prediction on a particular feature.

$L_1$  → Takes the absolute value of error.

$L_2$  Takes the squared.

Standard weight update

$$W \leftarrow W - \alpha \nabla_W f(W)$$

Taking account of Regu.

$$W \leftarrow W - \alpha \nabla_W f(W) + \lambda R(W)$$

↓

$$\alpha = \frac{1}{N} \sum \frac{1}{L_2} + \lambda R(W)$$

Hyperparameter

↓  
Reg. Penalty