Lecture 12

Name: Harsh Sanjay Roniyar

Roll Number: 22B3942

Continuing from where we left in Lecture 11,

$$\frac{dL}{dw} = \frac{dL}{dy}\frac{dy}{dh}\frac{dh}{dw};$$

where, $h = wx$, $y = \sigma(h) = \frac{1}{1+e^{-h}}$ and $L = -t\log(y) - (1-t)\log(1-y)$
Therefore, $\frac{dL}{dy} = \left(-\frac{t}{y} + \frac{1-t}{1-y}\right)$, $\frac{dy}{dh} = y(1-y)$.
Hence, we get

$$\frac{dL}{dw} = (y-t)x.$$

In linear as well as logistic regression, the model is still linear and loss functions are also convex.

Now, considering regularization, for L2 Regularization, the total loss comes out to be $L_{\text{error}} + \frac{\lambda}{2}w^T w$, whereas in L1 regularization, it is $L_{\text{error}} + \lambda\sum_j |w_j|$.
Also, L2 regularization has some sort of weight stabilization effect, i.e. it tries to correct the deviations of w_i's such that the loss is minimized, whereas L1 has no such effect.

Elastic Net

`Best of both L1 and L2 (but there's obviously a catch)`

$$\text{Total Loss } = L_{\text{error}} + \frac{\lambda_2}{2}\sum_j |w_j|^2 + \lambda_1 \sum_j |w_j|$$

The catch here is - Now we need to tune two hyperparameters instead of just one.

Confusion Matrix

Predicted \ Actual	0	1
0	TN	FN
1	FP	TP

Symmetric Risk

$(0, 0), (1, 1)$ - Low Risk

Predicted \ Actual	0	1
0	low	high
1	high	low

Asymmetric Risk

Different risks for **Type I (FP)** and **Type II (FN)** errors.

Predicted \ Actual	0	1
0	low	higher
1	high	low

Some metrics for binary classification

Single Threshold

$$\text{Accuracy} = \frac{TP + TN}{\text{Total}}$$

$$\text{Precision (P)} = \frac{TP}{TP + FP}$$

$$\text{Recall Sensitivity (R)} = \frac{TP}{TP + FN}$$

$$\text{Specificity} = \frac{TN}{TN + FP}$$

$$\text{Balanced Metric} : F1 \text{ score} = \frac{2PR}{P + R}$$

All the metrics defined above are threshold dependent, depending on the boundary condition set for classification between the classes 0 and 1, the values for all metrics will change accordingly.

Threshold-free Metric

AUC (Area Under (receiver operating characteristic) Curve)

The plot is made between sensitivity and $(1 \text{ - specificity})$. The farther the threshold point from the origin, the lower the threshold (variation of both values on the axes from 0 to 1).

```
Accuracy is meaningful only when we have balanced classes otherwise
preferable to use other metrics like AUC and F1-score.
```

Jensen's Inequality

Similar to what we did in Lecture 9.

A function f is convex iff $\forall\ 0 \leq \lambda \leq 1$ -

$$\lambda f(x_1) + (1-\lambda)f(x_2) \geq f(\lambda x_1 + (1-\lambda)x_2)$$