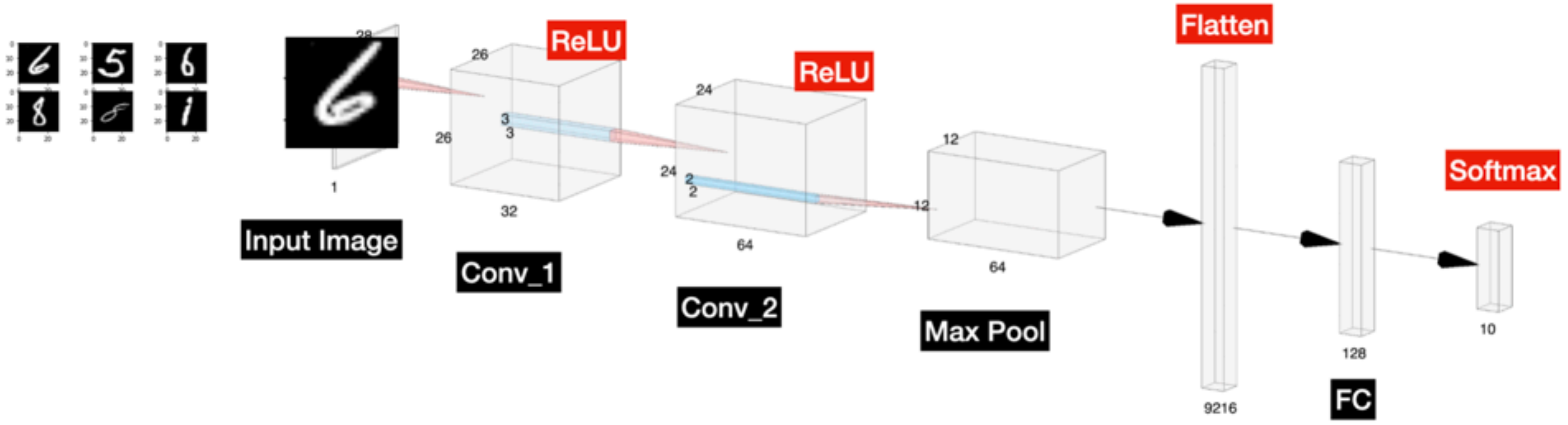
MODERN COMPUTER VISION

BY RAJEEV RATAN

Loss Functions

Loss Functions are essential to training

Quantifying Loss



P	6	5	6	8	8	1
0	0	0.9	0.83	0.21	0.19	0.62
1	0.73	0.8	0.89	0.7	0.92	0.07
2	0.78	0.88	0.19	0.39	0.08	0.74
3	0.37	0.56	0.07	0.64	0.64	0.9
4	0.63	0.25	0.79	0.94	0.52	0.55
5	0.87	0.65	0.57	0.63	0.97	0.04
6	0.67	0.05	0.45	0.51	0.87	0.51
7	0.71	0.66	0.13	0.59	0.86	0.89
8	0.51	0.88	0.59	0.01	0.37	0.63
9	0.24	0.52	0.79	0.15	0.63	0.78

- How bad are the probabilities we predicted?
- How do we quantify the degree our prediction is off by?

Cross Entropy Loss or Categorical Cross Entropy Loss

Class	Predicted Probabilities	Ground Truth
0	0.1	0
1	0.2	0
2	0.1	0
3	0.05	0
4	0.05	0
5	0.05	0
6	0.05	0
7	0.3	1
8	0.05	0
9	0.05	0

- Cross Entropy Loss uses two distributions, our ground truth distribution $p(x)$ and $q(x)$ our predicted distribution.
- $L = -y \, . \, log(\hat{y})$
- Where y is the ground truth vector, $\hat{y}$ is the predicted distribution and ‘ . ‘ is the inner product.

Cross Entropy Loss a Simpler Example

Class	Predicted Probabilities	Ground Truth
0	0.3	0
1	0.6	1
2	0.1	0

- $L = -y \, . \, log(\hat{y})$
- $L = -(0 \times log(0.3) + 1 \times log(0.6) + 0 \times log(0.1)$
- $L = -(0 + 1 \times -0.222 + 0) = 0.222$
- **NOTE:**
    - Multi-class log loss rewards/penalises the correct classes only

$$L = \frac{1}{N} \sum_{i-1}^{N} L_i$$

-

Other Loss Functions

- Loss Functions are sometimes called **Cost Functions**
- For Binary Classification problems we use **Binary Cross-Entropy Loss** (same as categorical cross-entropy loss except it uses just one output node)
- For Regressions we often use the Mean Square Error (MSE)
  - Mean Square Error (MSE) = (Target - Predicted)^2
  - $MSE = \frac{1}{n} \sum (Y_i - \hat{Y}_i)^2$
- Other loss functions that are sometimes used:
  - L1, L2
  - Hinge Loss
  - Mean Absolute Error (MAE)

What do we do with our Quantified Loss?

- Updating all the weights of our model is not trivial
- How do we correctly update our weights to minimise loss?
- We use Back Propagation
- And we use the loss value for this!

MODERN COMPUTER VISION

BY RAJEEV RATAN

Next…

Back Propagation