# CME 216, ME 343 - Spring 2020

Eric Darve, ICME

Let us review the loss functions that are available in TensorFlow.

See Keras loss documentation for the complete list.

`BinaryCrossentropy`

Cross entropy when only two labels (0 and 1) are possible.

`CategoricalCrossentropy`

Cross entropy with more than 2 labels.

This is the case we explained.

## Hinge

This is a variant which applies when the labels are either -1 or 1.

The formula is

loss = $\mathbf{max}(1 - y_{\mathbf{true}}\ y_{\mathbf{pred}}, 0)$

where $y_{\mathbf{true}}$ and $y_{\mathbf{pred}}$ are the labels.

`CategoricalHinge`

This is the hinge loss for the case where more than 2 labels are possible.

The formula is

loss = $\mathbf{max}(1 - \mathbf{pos} + \mathbf{neg}, 0)$

where:
pos = $\sum_i [y_{\text{true}}]_i \ [y_{\text{pred}}]_i$
neg = $\mathbf{max}_i((1 - [y_{\text{true}}]_i)[y_{\text{pred}}]_i)$

This applies when $y$ is a probability.

Let's assume that $\boldsymbol{y}_{\mathbf{true}}$ is a one-hot vector.

Then

pos = $\sum_i [\boldsymbol{y}_{\mathbf{true}}]_i \ [\boldsymbol{y}_{\mathbf{pred}}]_i = [\boldsymbol{y}_{\mathbf{pred}}]_t$

is the value of $\boldsymbol{y}_{\mathbf{pred}}$ for the true label $\boldsymbol{t}$.

neg = $\mathbf{max}_{i \neq t} [y_{\text{pred}}]_i$

is the maximum of $y_{\text{pred}}$ for all labels different from the true one $t$.

So, using compact notations

loss = $\mathbf{max}(1 - [y_{\text{pred}}]_t + \mathbf{max}_i [y_{\text{pred}}]_{i \neq t}, 0)$

For a good model:

$[y_{\text{pred}}]_t \approx 1$

$\max_i [y_{\text{pred}}]_{i \neq t} \ll 1$

So the loss will be small.

Note that the documentation for CategoricalHinge has a rather mistifying error.

Can you find it?

The formula for `neg` and `pos` is not correct for some reason.

Fortunately, the code is correct.