Deep Learning course

Session 9 – Loss functions

E. Francisco Roman-Rangel
edgar.roman@alumni.epfl.ch

CInC-UAEM. Cuernavaca, Mexico. September 29th, 2018.

Outline

Mean squared error

$$\mathcal{L}_{MSE} = \frac{1}{m}\sum_{i=1}^{m}(y^{(i)} - \hat{y}^{(i)})^2$$

Mean absolute error

$$\mathcal{L}_{MAE} = \frac{1}{m} \sum_{i=1}^{m} |y^{(i)} - \hat{y}^{(i)}|$$

Hinge loss

Classification (SVM-like):

$$\mathcal{L}_H = \max(0, 1 - y \cdot \hat{y})$$

Background: Entropy

Expected number of bits need to encode some information.

$$H(y) = \sum_i y_i \log \frac{1}{y_i} = -\sum_i y_i \log y_i$$

Categorical Cross Entropy

Number of bits needed if information from y is encoded using $\hat{y}$.

$$\mathcal{L}_{CE} = H(y, \hat{y}) = \sum_i y_i \log \frac{1}{\hat{y}_i}$$

Binary Cross Entropy (think of probabilities):

$$\mathcal{L}_{BCE} = -(y \log \hat{y} + (1 - y) \log(1 - \hat{y}))$$

Kullback–Leibler divergence

Number of extra bits needed on average if information from y is encoded according to $\hat{y}$.

$$\mathcal{L}_{D_{KL}} = \sum_i y_i \log \frac{y_i}{\hat{y}_i}$$

It's equivalent to categorical cross entropy up to a scaling factor.

Thank you.

Q&A