2020 資格スクエア
日本ディープラーニング協会 G 検定対策ビデオ
ディープラーニング 5-06
コスト関数

浅川伸一

コスト関数

- コスト関数 cost function
- 損失関数 loss function
- 誤差関数 error function
- 目的関数 objective function

$$p(\mathbf{y}|\mathbf{x};\theta) \tag{1}$$

最小二乗誤差（下式）, あるいは**負の対数尤度** negative log likelifood ($-\log(x)$) など

$$J(\theta) = \frac{1}{2}\mathbb{E}_{\mathbf{x},\mathbf{y}\sim\hat{p}_{data}}\|\mathbf{y} - f(\mathbf{x};\theta)\|^2 + \text{const.} \tag{2}$$

交差エントロピー損失関数

ニューラルネットワークや機械学習において，予測すべき値が2値化された量，たとえば真偽値真であれば 1 をとり，偽であれば 0 であったり，確率である場合には，最小化すべき目標関数(正則化項を含めて損失関数でもよい)は平均二乗誤差 Mean Square Errors ではなく **交差エントロピー cross-entropy 損失**，あるいは交差エントロピー誤差と呼ぶ関数が用いられる。

自乗誤差に比べて交差エントロピーを用いると学習が高速化される。 文献的にはニューラルネットワークに交差エントロピーが導入されたのは Hinton (1989)

交差エントロピーは次式で表される:

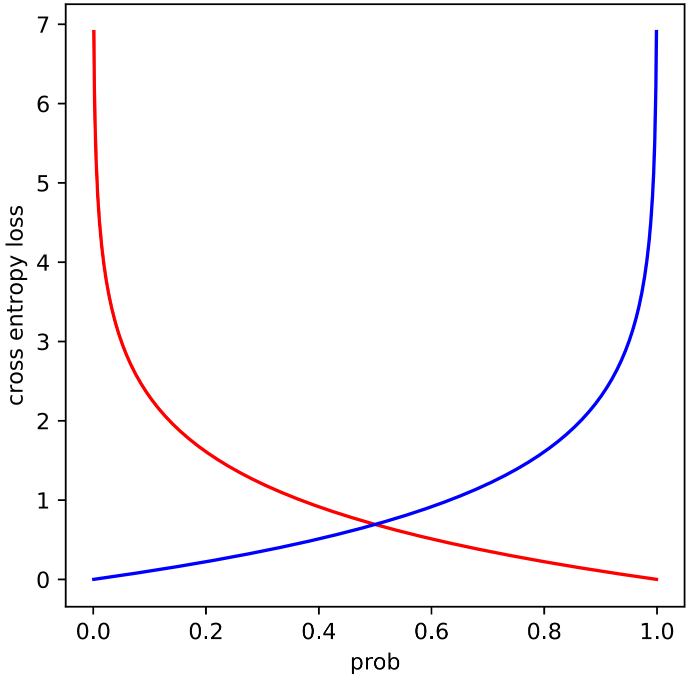
$$\mathcal{L} = -t\log(y) - (1-t)\log(1-y), \tag{3}$$

ここで t は教師信号すなわち 1 または 0 をとり，y はニューラルネットワークから出力された予測値。

上式は （確率とみなせる）出力 y が t 回起こった と解釈できる y^t このときの t の値はは 0 か 1 しか取らない ので，上式右辺は，もし t が 1 であれば右辺第一項を計算し，t が 0 であれば 右辺第2項を計算することになる。

右辺第一項と右辺第二項とを別曲線として描いた次ページ図。

交差エントロピーのグラフ



交差エントロピーのグラフ

ここで対数 $\log$ の底は e や 2 が用いられる。

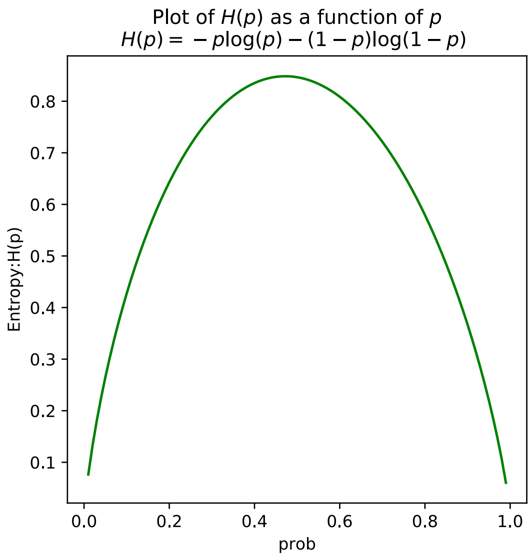
エントロピー

エントロピーには熱力学エントロピーと情報論的エントロピーと 2 種類存在するがどちらも同じ形式をしている。情報論的には平均エントロピー H を以下のように定義する

$$H[X] = -\sum_{i} X_i \log(X_i) \tag{4}$$

上式 は 平均情報量 (Shannon 1948) とも呼ばれる。連続変量の場合には総和記号 $\sum$ が積分記号 $\int$ となって

$$H[x] = -\int x \log(x)\ dx \tag{5}$$



シャノンのエントロピー

まとめ

- コスト関数，損失関数，誤差関数，目的関数，はほぼ同じような意味で用いられる
- 代表的なコスト関数として，最小自乗誤差，交差エントロピー誤差，などがある
- 出力が確率で与えられるような問題，たとえば，分類問題などでは交差エントロピー誤差関数が用いられる

クイズ

$t\log(y) + (1-t)\log(1-y)$ を何と呼んだでしょうか

クイズの答え

$t \log(y) + (1 - t) \log(1 - y)$ を何と呼んだでしょうか

交差エントロピー（誤差）関数

文献

Hinton, Geoffrey E. 1989. “Connectionist Learning Procedures.” *Artificial Intelligence* 40: 185–234.

Shannon, Claude E. 1948. “A Mathematical Theory of Communication.” *The Bell System Technical Journal* 27 (3): 379–423.