HW

Entropy

Exercise 12.13.15 For any $p \in (0,1)$ consider the *binary entropy function*

$$H(p) = -p \ln p - (1-p)\ln(1-p).$$

(a) Show that $H(p)$ is the entropy associated with a Bernoulli random variable.
(b) Verify the following relation between the derivative of the binary entropy and the logit function:

$$\frac{dH(p)}{dp} = -\ln\left(\frac{p}{1-p}\right).$$

(a) $H(p) = -\sum_i p(x_i) \log p(x_i)$ $\qquad p(x=1) = p$

$p(x=0) = 1-p$

$= -\left(p \ln p + (1-p)\ln(1-p)\right)$

$= -p \ln p - (1-p)\ln(1-p)$

(b) $\frac{dH(p)}{dp} = -\ln p - 1 + \ln(1-p) + 1$

$= -\ln\left(\frac{p}{1-p}\right)$

Exercise 12.13.3 (a) Define the mutual information of X and Y, given Z as

$$I(X,Y|Z) = H(X|Z) + H(Y|Z) - H(X,Y|Z).$$

Show that $I(X,Y|Z) = D_{KL}[p(x,y,z)||p(x|z)p(y|z)]$.

(b) Show that for any three random variables X, Y, and Z, we have:

$$H(X|Z) + H(Y|Z) \geq H(X,Y|Z). \quad \because I \geq 0$$

When is the identity satisfied? when $p(x,y|z) = p(x|z)p(y|z)$

X, Y Conditional independent given Z

(a) $I(X,Y|Z) = H(X|Z) + H(Y|Z) - H(X,Y|Z)$

$$= -\iint p(x,z) \ln p(x|z)\, dx\, dz$$

$$-\iint p(y,z) \ln p(y|z)\, dy\, dz$$

$$+\iiint p(x,y,z) \ln p(x,y|z)\, dx\, dy\, dz$$

$$= -\iiint p(x,y,z) \ln p(x|z)\, dx\, dy\, dz$$

$$-\iiint p(x,y,z) \ln p(y|z)\, dx\, dy\, dz$$

$$+\iiint p(x,y,z) \ln p(x,y|z)\, dx\, dy\, dz$$

$$= \iiint p(x,y,z) \ln \frac{p(x,y|z)}{p(x|z)p(y|z)}\, dx\, dy\, dz$$

$$= \iiint p(x,y,z) \ln \frac{p(x,y,z)}{p(x|z)p(y|z)p(z)}\, dx\, dy\, dz$$

$$= D_{KL}[p(x,y,z) \,||\, p(x|z)p(y|z)p(z)]$$

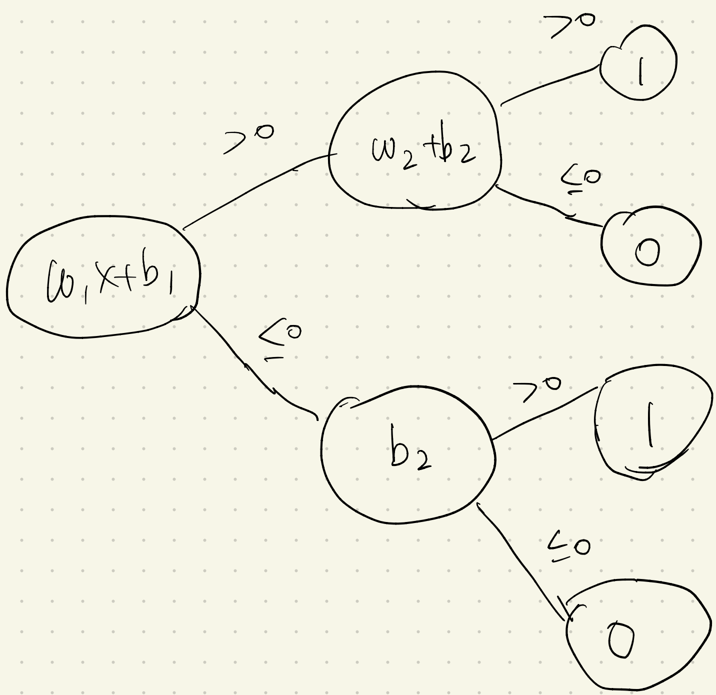
HW

Network Capacity

Exercise 12.13.7 Consider a neural network obtained by the concatenation of two perceptrons. The output of the network is given by the random variable

$$Y = H\big(w_2 H(w_1 X + b_1) + b_2\big),$$

with $X \in \{0, 1\}$. What is the capacity of this network? 2 bit



Exercise 12.13.9 How does the capacity of a network change when:
(a) An extra fully-connected layer is added to the network; 같거나 감소
(b) Some neurons are dropped out of the network; 감소
(c) The weights are constrained to be kept small. 감소

(a) layer 늘리면 at most Capacity 유지

(b), (c) : regularization
capacity를 줄이고 overfitting을 방지하여
보편적으로 useful한 모델이 되길 지향한다.