CS410: Artificial Intelligence 2021 Fall
Homework 5: Regression & Neural Networks & Bayes Nets
Due date: 23:59:59 (GMT +08:00), January 3 2022

1. **Cross entropy loss.** Recall the statement in Lecture 8, Slide 67 that the cross entropy loss function is convex in θ. Prove this statement in this exercise.

   Solution:

   Proof.

   Lemma 1. *A twice differentiable function of several variables is convex on a convex set if and only if its Hessian matrix of second partial derivatives is positive semidefinite on the interior of the convex set.*

   The Hessian matrix of the cross entropy loss function with respect to θ is

$$\begin{aligned}&\nabla_\theta^2 \mathcal{L}(y, x, p_\theta)\\ =&\frac{\partial}{\partial\theta}\left(\frac{\partial \mathcal{L}(y, x, p_\theta)}{\partial\theta}\right)\\ =&\frac{\partial}{\partial\theta}((\sigma(\theta^\top x) - y)x)\\ =&\sigma(\theta^\top x)(1-\sigma(\theta^\top x))xx^\top ,\end{aligned}$$

   which is positive semidefinite since $0 < \sigma(\theta^\top x) < 1$ and $\forall y \in \mathbb{R}^d$, $y^\top (xx^\top)y = (y^\top x)(x^\top y) = (y^\top x)^2 \geq 0$. The proof is concluded by Lemma 1. □

2. **Backpropagation.**

   (a) Calculate the output values at nodes h_1, h_2 and $\hat{y}$ of this network for input $x_1 = 0, x_2 = 1$. Show all steps in your calculation. Assume that the neurons have sigmoid activation function.

   (b) Compute one step of the backpropagation algorithm with $\eta = 1$ for a given example with input $x_1 = 0, x_2 = 1$ and target output $y = 1$, using new weights and old weights respectively. Compute the updated weights for both the hidden layer and the output layer. Comment on whether a further forward pass gives a lower error. Show all steps in your calculation. The error on the given example is defined as $E = 1/2(y - O)^2$ where O is the real-valued network output of that example at the output node, and y is the integer-valued target output for that example.

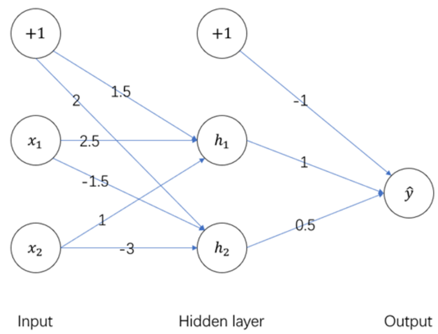


Figure 1: Problem 2.

Solution: Following the procedures in Slides 45-46 in Lecture 9 suffices. Final results may depend on rounding.

(a) Output of h_1 is $\sigma(1.5 \times 1 + 2.5 \times 0 + 1 \times 1) = 0.924$. Output of h_2 is $h_2 = \sigma(2 \times 1 - 1.5 \times 0 - 3 \times 1) = 0.269$. Output of $\hat{y}$ is $\hat{y} = \sigma(-1 \times 1 + 1 \times 0.924 + 0.5 \times 0.269) = 0.515$.

(b) New weights:

i. Output error: $\delta = (y - O)(1 - O)O = 0.1212$.

ii. Updated weights for output layer:

$$w_1^+ = -1 + (0.1212 \times 1) = -0.8788$$
$$w_2^+ = 1 + (0.1212 \times 0.9241) = 1.112$$
$$w_3^+ = 0.5 + (0.1212 \times 0.2689) = 0.5326$$

iii. Hidden layer error:

$$\delta_1 = 0.1212 \times 1.112 \times (1 - 0.9241) \times 0.9241 = 9.453 \times 10^{-3}$$
$$\delta_2 = 0.1212 \times 0.5326 \times (1 - 0.2689) \times 0.2689 = 1.269 \times 10^{-2}$$

iv. Updated weights for hidden layer:

$$w_4^+ = 1.5 + \left(9.453 \times 10^{-3} \times 1\right) = 1.509453$$
$$w_5^+ = 2.5 + \left(9.453 \times 10^{-3} \times 0\right) = 2.5$$
$$w_6^+ = 1 + \left(9.453 \times 10^{-3} \times 1\right) = 1.009453$$
$$w_7^+ = 2 + \left(1.269 \times 10^{-2} \times 1\right) = 2.01269$$
$$w_8^+ = -1.5 + \left(1.269 \times 10^{-2} \times 0\right) = -1.5$$
$$w_9^+ = -3 + \left(1.269 \times 10^{-2} \times 1\right) = -2.98731$$

v. New error: 0.0910. The value is reduced.

(c) Old weights:

i. Output error: 0.1212

ii. Hidden layer error:

$$\delta_1 = 0.1212 \times 1 \times (1 - 0.9241)0.9241 = 8.501 \times 10^{-3}$$
$$\delta_2 = 0.1212 \times 0.5 \times (1 - 0.2689)0.2689 = 1.191 \times 10^{-2}$$

iii. Updated weights:

$$w_1^+ = -1 + (0.1212 \times 1) = -0.8788$$
$$w_2^+ = 1 + (0.1212 \times 0.9241) = 1.112$$
$$w_3^+ = 0.5 + (0.1212 \times 0.2689) = 0.5326$$
$$w_4^+ = 1.5 + \left(8.501 \times 10^{-3} \times 1\right) = 1.508501$$
$$w_5^+ = 2.5 + \left(8.501 \times 10^{-3} \times 0\right) = 2.5$$
$$w_6^+ = 1 + \left(8.501 \times 10^{-3} \times 1\right) = 1.008501$$
$$w_7^+ = 2 + \left(1.191 \times 10^{-2} \times 1\right) = 2.01191$$
$$w_8^+ = -1.5 + \left(1.191 \times 10^{-2} \times 0\right) = -1.5$$
$$w_9^+ = -3 + \left(1.191 \times 10^{-2} \times 1\right) = -2.98809$$

iv. New error: 0.0910. The value is reduced.

3. **Bayes Nets.** Consider the following Bayes net. Calculate the marginal and conditional probabilities $\mathbb{P}(\neg P_3)$, $\mathbb{P}(P_2 \mid \neg P_3)$, $\mathbb{P}(P_1 \mid P_2, \neg P_3)$, and $\mathbb{P}(P_1 \mid \neg P_3, P_4)$ using **inference by enumeration**. Show all steps in your calculation.

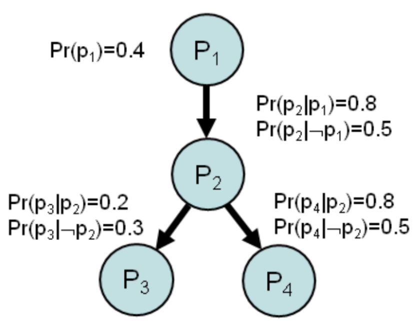


Figure 2: Problem 3.

Solutions:

$$
\begin{aligned}
&\mathbb{P}(\neg p_3)\\
&= \sum_{P_1,P_2,P_4} \mathbb{P}(P_1, P_2, \neg p_3, P_4)\\
&= \sum_{P_1,P_2,P_4} \mathbb{P}(P_1)\,\mathbb{P}(P_2 \mid P_1)\,\mathbb{P}(\neg p_3 \mid P_2)\,\mathbb{P}(P_4 \mid P_2)\\
&= \mathbb{P}(p_1)\,\mathbb{P}(p_2 \mid p_1)\,\mathbb{P}(\neg p_3 \mid p_2)\,\mathbb{P}(p_4 \mid p_2)\\
&+ \mathbb{P}(p_1)\,\mathbb{P}(p_2 \mid p_1)\,\mathbb{P}(\neg p_3 \mid p_2)\,\mathbb{P}(\neg p_4 \mid p_2)\\
&+ \mathbb{P}(p_1)\,\mathbb{P}(\neg p_2 \mid p_1)\,\mathbb{P}(\neg p_3 \mid \neg p_2)\,\mathbb{P}(p_4 \mid \neg p_2)\\
&+ \mathbb{P}(p_1)\,\mathbb{P}(\neg p_2 \mid p_1)\,\mathbb{P}(\neg p_3 \mid \neg p_2)\,\mathbb{P}(\neg p_4 \mid \neg p_2)+\\
&+ \mathbb{P}(\neg p_1)\,\mathbb{P}(p_2 \mid \neg p_1)\,\mathbb{P}(\neg p_3 \mid p_2)\,\mathbb{P}(p_4 \mid p_2)\\
&+ \mathbb{P}(\neg p_1)\,\mathbb{P}(p_2 \mid \neg p_1)\,\mathbb{P}(\neg p_3 \mid p_2)\,\mathbb{P}(\neg p_4 \mid p_2)+\\
&+ \mathbb{P}(\neg p_1)\,\mathbb{P}(\neg p_2 \mid \neg p_1)\,\mathbb{P}(\neg p_3 \mid \neg p_2)\,\mathbb{P}(p_4 \mid \neg p_2)\\
&+ \mathbb{P}(\neg p_1)\,\mathbb{P}(\neg p_2 \mid \neg p_1)\,\mathbb{P}(\neg p_3 \mid \neg p_2)\,\mathbb{P}(\neg p_4 \mid \neg p_2)\\
&= .4 \times .8 \times .8 \times .8 + .4 \times .8 \times .8 \times .2 + .4 \times .2 \times .7 \times .5 + .4 \times .2 \times .7 \times .5+\\
&+ .6 \times .5 \times .8 \times .8 + .6 \times .5 \times .8 \times .2 + .6 \times .5 \times .7 \times .5 + .6 \times .5 \times .7 \times .5 =\\
&= .2048 + .0512 + .028 + .028 + .192 + .048 + .105 + .105 = .762
\end{aligned}
$$

$$
\begin{aligned}
&\mathbb{P}(p_2 \mid \neg p_3)\\
&= \frac{\mathbb{P}(p_2, \neg p_3)}{\mathbb{P}(\neg p_3)} = \frac{.496}{.762} = .6509\\
&\mathbb{P}(p_2, \neg p_3)\\
&= \sum_{P_1,P_4} \mathbb{P}(P_1, p_2, \neg p_3, P_4) = \sum_{P_1,P_4} \mathbb{P}(P_1)\,\mathbb{P}(p_2 \mid P_1)\,\mathbb{P}(\neg p_3 \mid p_2)\,\mathbb{P}(P_4 \mid p_2)\\
&= \mathbb{P}(p_1)\,\mathbb{P}(p_2 \mid p_1)\,\mathbb{P}(\neg p_3 \mid p_2)\,\mathbb{P}(p_4 \mid p_2)\\
&+ \mathbb{P}(p_1)\,\mathbb{P}(p_2 \mid p_1)\,\mathbb{P}(\neg p_3 \mid p_2)\,\mathbb{P}(\neg p_4 \mid p_2)+\\
&+ \mathbb{P}(\neg p_1)\,\mathbb{P}(p_2 \mid \neg p_1)\,\mathbb{P}(\neg p_3 \mid p_2)\,\mathbb{P}(p_4 \mid p_2)\\
&+ \mathbb{P}(\neg p_1)\,\mathbb{P}(p_2 \mid \neg p_1)\,\mathbb{P}(\neg p_3 \mid p_2)\,\mathbb{P}(\neg p_4 \mid p_2)\\
&= .4 \times .8 \times .8 \times .8 + .4 \times .8 \times .8 \times .2 + .6 \times .5 \times .8 \times .8 + .6 \times .5 \times .8 \times .2\\
&= .2048 + .0512 + .192 + .048 = .496
\end{aligned}
$$

$\mathbb{P}(p_1 \mid p_2, \neg p_3)$
$= \dfrac{\mathbb{P}(p_1, p_2, \neg p_3)}{\mathbb{P}(p_2, \neg p_3)} = \dfrac{.256}{.496} = .5161$
$\mathbb{P}(p_1, p_2, \neg p_3)$
$= \sum_{P_4} \mathbb{P}(p_1, p_2, \neg p_3, P_4)$
$= \sum_{P_4} \mathbb{P}(p_1)\,\mathbb{P}(p_2 \mid p_1)\,\mathbb{P}(\neg p_3 \mid p_2)\,\mathbb{P}(P_4 \mid p_2)$
$= \mathbb{P}(p_1)\,\mathbb{P}(p_2 \mid p_1)\,\mathbb{P}(\neg p_3 \mid p_2)\,\mathbb{P}(p_4 \mid p_2)$
$+ \mathbb{P}(p_1)\,\mathbb{P}(p_2 \mid p_1)\,\mathbb{P}(\neg p_3 \mid p_2)\,\mathbb{P}(\neg p_4 \mid p_2)$
$= .4 \times .8 \times .8 \times .8 + .4 \times .8 \times .8 \times .2 = .2048 + .0512 = .256$
$\mathbb{P}(p_2, \neg p_3)$
$= \mathbb{P}(p_1, p_2, \neg p_3) + \mathbb{P}(\neg p_1, p_2, \neg p_3) = .256 + .24 = .496$
$\mathbb{P}(\neg p_1, p_2, \neg p_3)$
$= \sum_{P_4} \mathbb{P}(\neg p_1, p_2, \neg p_3, P_4) = \sum_{P_4} \mathbb{P}(\neg p_1)\,\mathbb{P}(p_2 \mid \neg p_1)\,\mathbb{P}(\neg p_3 \mid \mathrm{P}_2)\,\mathbb{P}(p_4 \mid P_2)$
$= \mathbb{P}(\neg p_1)\,\mathbb{P}(p_2 \mid \neg p_1)\,\mathbb{P}(\neg p_3 \mid p_2)\,\mathbb{P}(p_4 \mid p_2)$
$+ \mathbb{P}(\neg p_1)\,\mathbb{P}(p_2 \mid \neg p_1)\,\mathbb{P}(\neg p_3 \mid p_2)\,\mathbb{P}(\neg p_4 \mid p_2)$
$= .6 \times .5 \times .8 \times .8 + .6 \times .5 \times .7 \times .2 = .192 + .048 = .24$

$$\mathbb{P}(p_1 \mid \neg p_3, p_4)$$
$$= \frac{\mathbb{P}(p_1, \neg p_3, p_4)}{\mathbb{P}(\neg p_3, p_4)} = \frac{.2328}{.5298} = .4394$$
$$\mathbb{P}(p_1, \neg p_3, p_4)$$
$$= \sum_{P_2} \mathbb{P}(p_1, P_2, \neg p_3, p_4) = \sum_{P_2} \mathbb{P}(p_1)\,\mathbb{P}(P_2 \mid p_1)\,\mathbb{P}(\neg p_3 \mid \mathrm{P}_2)\,\mathbb{P}(p_4 \mid \mathrm{P}_2)$$
$$= \mathbb{P}(p_1)\,\mathbb{P}(p_2 \mid p_1)\,\mathbb{P}(\neg p_3 \mid p_2)\,\mathbb{P}(p_4 \mid p_2)$$
$$+ \mathbb{P}(p_1)\,\mathbb{P}(\neg p_2 \mid p_1)\,\mathbb{P}(\neg p_3 \mid \neg p_2)\,\mathbb{P}(p_4 \mid \neg p_2)$$
$$= .4 \times .8 \times .8 \times .8 + .4 \times .2 \times .7 \times .5 = .2048 + .028 = .2328$$
$$\mathbb{P}(\neg p_3, p_4)$$
$$= \mathbb{P}(p_1, \neg p_3, p_4) + \mathbb{P}(\neg p_1, \neg p_3, p_4) = .2328 + .297 = .5298$$
$$\mathbb{P}(\neg p_1, \neg p_3, p_4)$$
$$= \sum_{P_2} \mathbb{P}(\neg p_1, P_2, \neg p_3, p_4) = \sum_{P_2} \mathbb{P}(\neg p_1)\,\mathbb{P}(P_2 \mid \neg p_1)\,\mathbb{P}(\neg p_3 \mid P_2)\,\mathbb{P}(p_4 \mid P_2)$$
$$= \mathbb{P}(\neg p_1)\,\mathbb{P}(p_2 \mid \neg p_1)\,\mathbb{P}(\neg p_3 \mid p_2)\,\mathbb{P}(p_4 \mid p_2)$$
$$+ \mathbb{P}(\neg p_1)\,\mathbb{P}(\neg p_2 \mid \neg p_1)\,\mathbb{P}(\neg p_3 \mid \neg p_2)\,\mathbb{P}(p_4 \mid \neg p_2)$$
$$= .6 \times .5 \times .8 \times .8 + .6 \times .5 \times .7 \times .5 = .192 + .105 = .297$$

4. **Bayes Nets.** Consider the same Bayes net in Exercise 3. Compute $\mathbb{P}(\neg P_3)$ and $\mathbb{P}(P_2 \mid \neg P_3)$ using **variable elimination**. Compare the computational complexity of inference by enumeration and variable elimination and discuss your findings. Show all steps in your calculation.

   Solutions:

$$f_1(P_2) = \sum_{P_1} \mathbb{P}(P_1, P_2) = \sum_{P_1} \mathbb{P}(P_1)\,\mathbb{P}(P_2 \mid P_1)$$
$$f_1(p_2) = .4 \times .8 + .6 \times .5 = .62,\ f_1(\neg p_2) = .4 \times .2 + .6 \times .5 = .38$$

$$f_{1,2}(P_3) = \sum_{P_2} f_1(P_2)\,\mathbb{P}(P_3 \mid P_2)$$
$$f_{1,2}(p_3) = .62 \times .2 + .38 \times .3 = .238,\ f_{1,2}(\neg p_3) = .62 \times .8 + .38 \times .7 = .762$$

$$f_{1,3}(P_2) = f_1(P_2)\,\mathbb{P}(\neg p_3 \mid P_2)$$
$$f_{1,3}(p_2) = .62 \times .8 = .496,\ f_{1,3}(\neg p_2) = .38 \times .7 = .266$$

$$\mathbb{P}(\neg p_3) = f_{1,2}(\neg p_3) = .762$$
$$\mathbb{P}(p_2 \mid \neg p_3) = \frac{f_{1,3}(p_2)}{f_{1,3}(p_2) + f_{1,3}(\neg p_2)} = \frac{.496}{.496 + .266} = .6509$$

5. **Independence.** Answer the following questions by explicitly showing all steps in your calculation.

   (a) Is D independent from A given B in Figure 3?

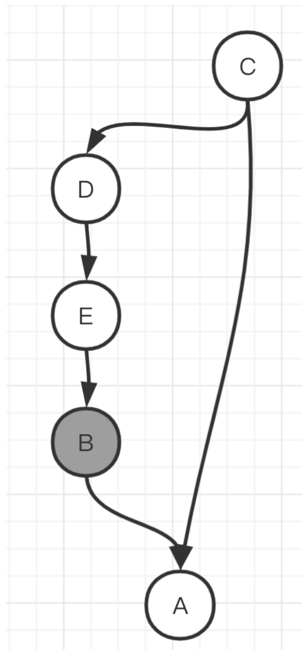


Figure 3: Problem 5.1.

   (b) Is D independent from C given E in Figure 4?

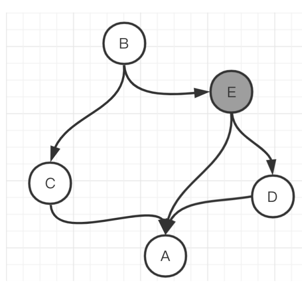


Figure 4: Problem 5.2.

   (c) Is D independent from A given E in Figure 5?

Solutions:

(a) No. Path $D \to C \to A$ is active.

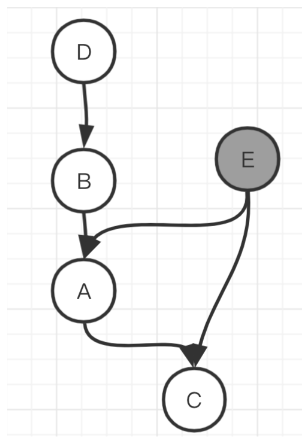
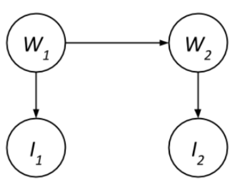


Figure 5: Problem 5.3.

(b) Yes. Path $C \to B \to E \to D$ is inactive. Path $C \to A \to D$ is inactive. Path $C \to B \to E \to A \to D$ is inactive. path $C \to A \to E \to D$ is inactive.

(c) No. Path $D \to B \to A$ is active.

6. **Likelihood Weighting.** Consider the following Bayesian network and the corresponding probabilities. Assume we generate the following six samples given the evidence $I_1 = T$ and $I_2 = F$: $(W_1, I_1, W_2, I_2) = \{(S, T, R, F), (R, T, R, F), (S, T, R, F), (S, T, S, F), (S, T, S, F), (R, T, S, F)\}$



W_1	$P(W_1)$
S	0.6
R	0.4

W_1	W_2	$P(W_2 \mid W_1)$
S	S	0.7
S	R	0.3
R	S	0.5
R	R	0.5

W	I	$P(I \mid W)$
S	T	0.9
S	F	0.1
R	T	0.2
R	F	0.8

Figure 6: Problem 6.

(a) What is the weight of the first sample (S, T, R, F) above?

(b) Use likelihood weighting to estimate $P(W_2 | I_1 = T, I_2 = F)$.

Solutions:

(a) The evidence is $I_1 = \mathrm{T}, I_2 = \mathrm{F}$.

$$w = \Pr(I_1 = \mathrm{T} \mid W_1 = \mathrm{S}) \cdot \Pr(I_2 = \mathrm{F} \mid W_2 = \mathrm{R}) = 0.9 \cdot 0.8 = 0.72.$$

(b) The sample weights are given by:

(W_1, I_1, W_2, I_2)	w	(W_1, I_1, W_2, I_2)	w
S, T, R, F	0.72	S, T, S, F	0.09
R, T, R, F	0.16	S, T, S, F	0.09
S, T, R, F	0.72	R, T, S, F	0.02

Therefore, we have

$$\hat{P}\left(W_2 = \mathrm{R} \mid I_1 = \mathrm{T}, I_2 = \mathrm{F}\right) = \frac{0.72 + 0.16 + 0.72}{0.72 + 0.16 + 0.72 + 0.09 + 0.09 + 0.02} = 0.889$$

$$\hat{P}\left(W_2 = \mathrm{S} \mid I_1 = \mathrm{T}, I_2 = \mathrm{F}\right) = 1 - 0.889 = 0.111$$