HW2

February 18, 2022

Problem 1 [Logistic regression - 2pts] Complete the following derivation: find the derivatives of the loss term w.r.t. $\boldsymbol{\theta}$:

Given the loss term

$$\ell(\boldsymbol{\theta}) = y \log(h(\boldsymbol{\theta}^\top \boldsymbol{x})) + (1-y)\log(1 - h(\boldsymbol{\theta}^\top \boldsymbol{x}))$$

where

$$h(z) = \frac{1}{1+e^{-z}}, \quad \boldsymbol{\theta} = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \end{bmatrix}, \quad \text{and} \quad \boldsymbol{x} = \begin{bmatrix} 1 \\ x_1 \\ x_2, \end{bmatrix},$$

write down the analytical expression of the derivative of $\ell(\boldsymbol{\theta})$ w.r.t. $\boldsymbol{\theta}$:

$$\nabla_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\partial \ell(\boldsymbol{\theta})}{\partial \theta_0} \\ \frac{\partial \ell(\boldsymbol{\theta})}{\partial \theta_1} \\ \frac{\partial \ell(\boldsymbol{\theta})}{\partial \theta_2} \end{bmatrix} =$$

Write down a step-by-step derivation. You will have to use the following mathematical facts. In every step of the derivation, explain how you used these mathematical facts.

- use the chain-rule: $\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{\mathrm{d}y}{\mathrm{d}z}\frac{\mathrm{d}z}{\mathrm{d}x}$
- derivative of the log function: $\frac{\mathrm{d}\log(x)}{\mathrm{x}} = \frac{1}{x}$
- derivative of $h(x) = \frac{1}{1+e^{-x}}$: $\frac{\mathrm{d}h(x)}{\mathrm{x}} = h(x)(1-h(x))$

Problem 2 [GLM and softmax regression - 3pts] In the lecture, we saw that linear regression and logistic regression can be derived as generalized linear models (GLMs) when we model the conditional probability $p(y|x)$ as Gaussian distribution and Bernoulli distribution, respectively.

In this homework, you will be asked to derive softmax regression from categorical distribution, $p(y|x) \sim \text{Categorical}(\{\phi_k\}_{k=1}^K)$ (assuming K classes).

Exponential family First, confirm that the categorical distribution $p(y)$ is an exponential family. Let's start by parameterizing the probabilities:

$$
\begin{aligned}
p(y=1) &= \phi_1, \\
p(y=2) &= \phi_2, \\
&\vdots \\
p(y=K-1) &= \phi_{K-1}, \\
p(y=K) &= 1 - \sum_{j=1}^{K-1} \phi_j,
\end{aligned}
$$

which can be written as

$$
p(y) = \phi_1^{\mathbb{1}[y=1]} \phi_2^{\mathbb{1}[y=2]} \cdots \phi_{K-1}^{\mathbb{1}[y=K-1]} \phi_K^{\mathbb{1}[y=K]}.
$$

Write $p(y)$ in the form of the exponential family

$$
p(y) = b(y) \exp(\eta^\top T(y) - a(\eta))
$$

by filling in the blanks below (show your work):

$$
\begin{aligned}
b(\eta) &= \\
a(\eta) &= \\
\eta &= \begin{bmatrix} \\ \\ \\ \\ \\ \end{bmatrix} \in \mathbb{R}^{k-1}.
\end{aligned}
$$

For this, $T(y)$ should be defined as a vector $T(y) \in \mathbb{R}^{K-1}$ such that

$$
T(1) = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, T(2) = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \cdots T(k-1) = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}, T(k) = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{1}
$$

GLM Now, you will be finishing the derivation of the softmax regression from the conditional probability $p(y|x)$:

1. Assume that $p(y = 1|x) = \phi_i$ (Categorical distribution)
2. Using the definition of $T(y)$ as in Eq. (1), the parameterized model will be derived from the following equation:
$$h_\theta(x) = \mathbb{E}[T(y)|x; \theta]. \tag{2}$$
3. With an additionally assumption, $\eta = \theta^\mathsf{T} x$, applying some mathematical manipulation on Eq. (2) will yield
$$h_\theta(x) = \begin{bmatrix} \frac{\exp(\theta_1^\mathsf{T} x)}{\sum_{j=1}^K \exp(\theta_j^\mathsf{T} x)} \\ \frac{\exp(\theta_2^\mathsf{T} x)}{\sum_{j=1}^K \exp(\theta_j^\mathsf{T} x)} \\ \vdots \\ \frac{\exp(\theta_{K-1}^\mathsf{T} x)}{\sum_{j=1}^K \exp(\theta_j^\mathsf{T} x)} \end{bmatrix}. \tag{3}$$

Derive Eq. (3) from Eq. (2) by using the assumptions 1–3 and the exponential family expression of the categorical distribution. Show your work.

Problem 3 [Naïve Bayes - 2pts] Given the table below,

Day	Outlook	Wind	PlayTennis
D1	Sunny	Weak	Yes
D2	Sunny	Strong	No
D3	Overcast	Weak	Yes
D4	Rain	Strong	No
D5	Rain	Weak	Yes
D6	Rain	Strong	No
D7	Overcast	Strong	Yes
D8	Sunny	Weak	Yes

Table 1:

complete the following probability look-up tables:

Outlook	PlayTennis = Yes	PlayTennis = No
Sunny		
Overcast		
Rain		

Table 2: Probability look-up table

Wind	PlayTennis = Yes	PlayTennis = No
Weak		
Strong		

Table 3: Probability look-up table [1.0pt= 0.25 $\times$ 4]

Predict whether to play tennis if Outlook = Sunny and Wind = Strong.
Correct prediction: