Deep Learning: Homework #1

Due: February 6, 2026 at 11:59 pm
Total Points: 26

Instructions

The assignment below consists of the seven (7) questions, whose solutions can be awarded a maximum of 26 total points. Be sure to show your work as credit is given for intermediate solutions. You should submit your solutions to Gradescope as a PDF, using LaTeX to typset the mathematical expressions.

Count-Valued Regression

Let's consider building a generalized linear models (GLM) for count-valued data, meaning that the label's support is the non-negative integers, i.e. $\mathrm{y} \in \mathbb{N}_{\geq 0}$. One way to construct such a model is to assume the count value is conditionally drawn from a Poisson distribution:

$$\mathrm{y} \sim \text{Poisson}(\mathrm{y}; \lambda), \quad \text{Poisson}(\mathrm{y}; \lambda) \triangleq \frac{\lambda^{\mathrm{y}} \exp\{-\lambda\}}{\mathrm{y}!},$$

where $\lambda \in (0, \infty)$ is the rate parameter. The Poisson distribution expresses the probability that a given number of events occur in a fixed time interval if these events occur independently and with a constant mean rate. The rate parameter represents both the mean and variance: $\mathbb{E}[\mathrm{y}] = \text{Var}[\mathrm{y}] = \lambda$.

Another way to construct a model for counts is via the Geometric distribution:

$$\mathrm{y} \sim \text{Geometric}(\mathrm{y}; \pi), \quad \text{Geometric}(\mathrm{y}; \pi) \triangleq (1-\pi)^{\mathrm{y}} \cdot \pi,$$

where $\pi \in (0, 1]$ is known as the 'success probability.' The Geometric expresses the number of Bernoulli trials, each with chance of success

π, that fail before a success is reached. Its mean is $\mathbb{E}[\mathrm{y}] = (1-\pi)/\pi$, and its variance is $\mathrm{Var}[\mathrm{y}] = (1-\pi)/\pi^2$.

Now assume that we construct two GLMs, one for each of these distributional assumptions. Yet for both models, we will parameterize the expectation in the same way:

$$\mathbb{E}[\mathrm{y}|\mathrm{x}] \triangleq \exp\{\mathrm{w} \cdot \mathrm{x}\}, \tag{1}$$

where $\mathrm{x} \in \mathbb{R}$ is the feature (i.e. independent variable) and $\mathrm{w} \in \mathbb{R}$ is the (slope) parameter. This makes the link function $g(\mathrm{z}) = \log \mathrm{z}$ and the inverse link $g^{-1}(\mathrm{z}) = \exp\{\mathrm{z}\}$.

Problem 1

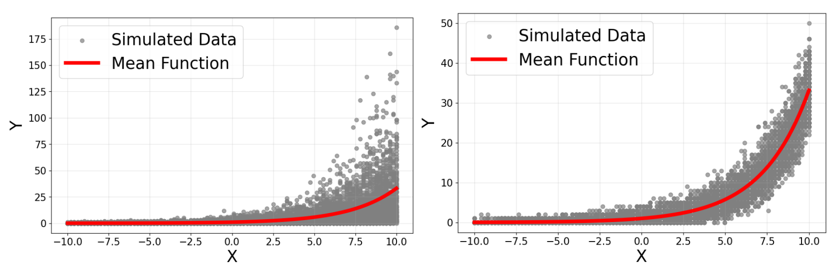
In Equation 1, assume $\mathrm{w} = 0.35$. For both the Poisson and Geometric GLMs defined above, calculate the mean and variance at $\mathrm{x} = 10$, i.e. $\mathbb{E}[\mathrm{y}|\mathrm{x} = 10]$ and $\mathrm{Var}[\mathrm{y}|\mathrm{x} = 10]$. (6 Points)

$w = 0.35$, $E[y|x = 10] = exp\{0.35 \times 10\} = e^{3.5} = 33.12$ for both Poisson and Geometric.

For Poisson GLM, $E(y) = Var(y)$, $\therefore Var(y|x = 10) = 33.12$

For Geometric GLM, $E(y) = \frac{1-\pi}{\pi}$, $Var(y) = \frac{1-\pi}{\pi^2}$

$\therefore Var(y) = E(y) \times (1 + E(y)) = 1129.75$



(a) Model #1

(b) Model #2

Figure 1: Two generalized linear models for count-valued data. One subfigure is the Geometric GLM and one is the Poisson GLM, but you must figure out which is which in problem #2. The red line shows the mean function $\mathbb{E}[\mathrm{y}|\mathrm{x}]$, and the gray dots show data that was sampled from each model.

Problem 2
Now examine the two models shown in the subfigures of Figure 1. Each model's parameter is set as w = 0.35, just as was assumed in problem #1. Which model is the Poisson GLM and which is the Geometric GLM? Justify your answers, possibly using your solution from problem #1. Hint: note the scales of the y-axes. (4 Points)

Model #1 is the Geometric GLM model and Model #2 is the Poisson GLM model.
From Problem #1, the two models have the same E, but the Geometric model's Variance is much bigger than Poisson model, so it will have much greater spread and more extreme values in y-axes. Model #1 (Figure a), its y-axis extends to around 180 and the simulated data show very large dispersion. So it shoule be the Geometric GLM model. And Model #2(Figure b) has a smaller spread in the scales of the y-axes.

Problem 3
Assume we observe a data set $\mathcal{D} = \{(x_n, y_n)\}_{n=1}^N$ and wish to fit the Poisson GLM to this data. Derive a loss function for the Poisson GLM using the framework of maximum likelihood estimation. Simplify the loss function as much as possible. Hint: see the derivation for logistic regression in Section 1.5 of the notes. (4 Points)

$L(w, D) = \prod_{n=1}^N \frac{\lambda^{y_n} exp\{-\lambda\}}{(y_n)!}$

Loss Function is the negative log-likelihood, so

$l(w, D) = -log(L(w, D)) = \sum_{n=1}^N [-log(\lambda^{y_n} exp\{-\lambda\}) + log(y_n!)] = \sum_{n=1}^N [-y_n log\lambda + \lambda + log(y_n!)]$

$\lambda = exp\{w \cdot x\}$, $\therefore l(w, D) = \sum_{n=1}^N [-y_n w \cdot x_n + exp\{w \cdot x_n\} + log(y_n!)]$

Problem 4
Using the final form of your loss function from problem #3, derive an expression for the derivative with respect to the regression parameter w; that is $\frac{d}{dw}\ell(w; \mathcal{D})$. (3 Points)

$l(w, D) = \sum_{n=1}^{N}[-y_n w \cdot x_n + exp\{w \cdot x_n\} + log(y_n!)]$

$\frac{d}{dw}l(w, D) = \sum_{n=1}^{N}[-x_n y_n + x_n exp\{w \cdot x_n\}]$

Problem 5
Now assume that we wish to fit the Geometric GLM to data. Derive a loss function for the Geometric GLM using the framework of maximum likelihood estimation. (4 Points)

$L(w, D) = \prod_{n=1}^{N}(1-\pi)^{y_n} \cdot \pi$

Loss Function is the negative log-likelihood, so

$l(w, D) = -log(L(w, D)) = \sum_{n=1}^{N}[-y_n log(1-\pi) - log\pi]$

$\frac{1-\pi}{\pi} = exp\{w \cdot x\}$, $\therefore \pi = \frac{1}{exp\{w \cdot x\}+1}$, $1 - \pi = \frac{exp\{w \cdot x\}}{exp\{w \cdot x\}+1}$

$\therefore log\pi = -log[exp\{w \cdot x\}+1]$, $log(1-\pi) = w \cdot x - log(exp\{w \cdot x\}+1)$

Then we get $l(w, D) = \sum_{n=1}^{N}[-y_n w \cdot x_n + y_n log(exp\{w \cdot x_n\} + 1) +$

$log(exp\{w \cdot x_n\} + 1)]$

$= \sum_{n=1}^{N}[-y_n w \cdot x_n + (y_n + 1)log(exp\{w \cdot x_n\} + 1)]$

Problem 6
Using the final form of your loss function from problem #5, derive an expression for the derivative with respect to the regression parameter w; that is $\frac{d}{dw}\ell(w; \mathcal{D})$. (3 Points)

$l(w, D) = \sum_{n=1}^{N}[-y_n w \cdot x_n + (y_n + 1)log(exp\{w \cdot x_n\} + 1)]$

$\frac{d}{dw}l(w, D) = \sum_{n=1}^{N}[-y_n x_n + (y_n + 1)x_n \frac{exp\{w \cdot x_n\}}{exp\{w \cdot x_n\}+1}]$

Problem 7

1. What similarities do you notice between the form of the Poisson GLM's and Geometric GLM's derivatives? (1 Point)
2. The derivatives computed in problems #4 and #6 should involve a sum over data points. Considering the derivative for just one data point (i.e. just one term in the sum), when will this single-data-point derivative be zero? Ignore the trivial case in which $x_n = 0$. (1 Point)

1.From the above problems. we get

$\nabla l(w, D)_{Poisson} = \sum_{n=1}^{N}[-x_n y_n + x_n exp\{w \cdot x_n\}]$

$\nabla l(w, D)_{Geometric} = \sum_{n=1}^{N}[-y_n x_n + (y_n + 1)x_n \frac{exp\{w \cdot x_n\}}{exp\{w \cdot x_n\}+1}]$

Both terms can be written as the form: $x_n \times$[model predicted data - observed data], where observed data is y_n and the predicted data is different.

2.Since $x_n \neq 0$, if the derivatives equal to zero, it will be:

$exp\{w \cdot x_n\} - y_n = 0$ and

$(y_n + 1)\frac{exp\{w \cdot x_n\}}{exp\{w \cdot x_n\}+1} - y_n = 0$

It means the model's predicted data exactly matches the observed one.