

CS480/680, Spring 2023

Assignment 1

Designer: Haochen Sun; Instructor: Hongyang Zhang

Released: May 15; Due: June 4, noon

Instructions

- We do not accept hand-written submissions.
- This assignment is due by noon on June 4, 2023.
- For questions labelled as “**coding**”, please follow the instructions provided and implement the required features. The skeleton code is provided. Unless otherwise specified, all implementations should be in *Python* using a *Jupyter Notebook*. Before submission, please make sure that your code can run without any errors. Also, be sure to save the output of each cell, as any missing output may not be graded.
- Please submit the following TWO files to LEARN:
 - A write-up in PDF format: the written answers to ALL questions, including the reported results and plots of coding questions, in a single PDF file.
 - An IPYNB file: your implementations for ALL coding questions. Please save the output of each cell, or your coding questions may NOT be graded.

Question 1: Perceptron (30 points)

1. [15 points] Consider the AND, OR and XOR datasets, each of which labels all two-dimensional binary inputs $\left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}$:

	$\begin{pmatrix} 0 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 0 \\ 1 \end{pmatrix}$	$\begin{pmatrix} 1 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 1 \\ 1 \end{pmatrix}$
AND	0	0	0	1
OR	0	1	1	1
XOR	0	1	1	0

Table 1: The AND, OR and XOR datasets

For each dataset, prove or disprove if it is linearly separable. For a linearly separable dataset, write down the separating hyperplane. For a non-linearly separable dataset, argue that a separating hyperplane does not exist.

2. [15 points, coding] Implement the perceptron algorithm on the Spambase dataset in *Python* using *Jupyter Notebook*. You may use the provided skeleton code, which downloads and pre-processes the dataset. Note that the target variable to be predicted is the last feature, `is_spam`. Plot the accuracy against the number of training steps and report the final accuracy.

Question 2: Generalized Linear Models (40 points)

In class, we have discussed linear regression and logistic regression. While these models are useful for specific types of data, they belong to a broader class of models known as generalized linear models (GLMs). GLMs are models for data where the response variable y follows a distribution from the exponential family, and the mean of the response variable is related to the predictors via a link function. The GLM has the following form:

$$p(y|\mathbf{x}, \mathbf{w}, \tau) = h(y, \tau) \exp \left[\frac{y\eta - A(\eta)}{d(\tau)} \right], \quad (1)$$

where p is the probability mass function or probability density function of y , $\eta := \mathbf{w}^\top \mathbf{x}$ is the natural parameter, $A(\eta)$ is the log normalizer, and τ is the dispersion parameter, such that $d(\tau) > 0$ typically related to the variance of the conditional distribution. $h(y, \tau)$ is a normalization factor such that $p(y|\mathbf{x}, \mathbf{w}, \tau)$ sums/integrates to 1 over y . We denote the mapping from the linear input $\eta = \mathbf{w}^\top \mathbf{x}$ to the conditional expectation of y (i.e., $\mu = \mathbb{E}[y|\mathbf{x}, \mathbf{w}, \tau]$) as $\mu = \ell^{-1}(\eta)$, where ℓ is known as the link function, and ℓ^{-1} is known as the mean function. (For simplicity, omit the bias term throughout this question.)

1. **[10 points]** Show that linear regression and logistic regression are special cases of the GLM in (1). Identify $h(y, \tau)$, $A(\eta)$, $d(\tau)$, and the link function ℓ for each case.
2. **[6 points]** Consider the Poisson regression model, which is defined by the following probability mass function:

$$p(y|\mathbf{x}, \mathbf{w}) = \text{Pois}(y|\exp(\mathbf{w}^\top \mathbf{x})).$$

Here, $\text{Pois}(y|\mu) = \frac{\mu^y e^{-\mu}}{y!}$ is the Poisson distribution parameterized by μ , with support $y \in \mathbb{N}$. Show that Poisson regression belongs to the family of GLMs defined in Equation (1), and identify $A(\eta)$ for this model.

3. **[8 points]** Consider a dataset $\{(\mathbf{x}_n, y_n)\}_{n=1}^N$, where $\mathbf{x}_n \in \mathbb{R}^d$ and y_n is the response variable. Derive the negative log-likelihood (NLL) for the GLM on this dataset based on Equation (1). Explain how the derived NLL is equivalent to the square loss (for linear regression) and binary cross entropy loss (for logistic regression) for finding the optimal \mathbf{w} .
4. **[6 points]** Derive the negative log-likelihood (NLL) loss function for the Poisson regression model using the results from sub-questions 2 and 3. Simplify the expression as much as possible.
5. **[10 points]** Prove that in Equation (1), $\mathbb{E}[y|\mathbf{x}, \mathbf{w}, \tau] = A'(\eta)$ and $\text{Var}[y|\mathbf{x}, \mathbf{w}, \tau] = A''(\eta)d(\tau)$. (Hint: you may either assume that y is discrete or continuous. As the first step, sum/integrate both sides of Equation (1) over y . You may switch the order of the summation/integration and the derivative without justification.)

Question 3: SVM kernels (30 points)

1. [10 points] Given a natural number M , consider $X = \{0, 1, \dots, M\}$. Define $K(x, x') = \min\{x, x'\}$. Find a mapping $\phi : X \rightarrow \mathbb{R}^M$ such that for all $x, x' \in X$, $K(x, x') = \langle \phi(x), \phi(x') \rangle$.

2. [10 points] Show that there exists a Hilbert space H and a mapping $\phi : \mathbb{R}^n \rightarrow H$ such that

$$\langle \phi(\mathbf{u}), \phi(\mathbf{v}) \rangle = 4 \langle \mathbf{u}, \mathbf{v} \rangle^2 + \langle \mathbf{u}, \mathbf{v} \rangle^3$$

for all $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$. (Hint: consider $H = \mathbb{R}^{n^2+n^3}$.)

3. [10 points] More generally, consider a polynomial f with non-negative coefficients, and construct H and ϕ such that

$$\langle \phi(\mathbf{u}), \phi(\mathbf{v}) \rangle = f(\langle \mathbf{u}, \mathbf{v} \rangle)$$

for all $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$.