HW Assignment 3, Theory

Problems marked with a (∗) are mandatory only for ITCS 8156 students. Bonus problems are optional, solving them will result in extra points.

1 Maximum Likelihood Estimation (20 points)

The Poisson distribution specifies the probability of observing k events in an interval, as follows:

$$P(k \text{ events in interval}) = e^{-\lambda} \frac{\lambda^k}{k!} \tag{1}$$

For example, k can be the number of meteors greater than 1 meter diameter that strike Earth in a year, or the number of patients arriving in an emergency room between 10 and 11 pm[1].

Suppose we observe N samples k_1, k_2, ..., k_N from this distribution (i.e. numbers of meteors that strike Earth over a period of N years). Derive the maximum likelihood estimate of the event rate λ.

2 Feature Scaling (30 points)

Consider the training and test examples shown in the first table below:

Original

	Train			Test
	$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$
ϕ_1	0	1	2	-1
ϕ_2	1	2	3	2
ϕ_3	2	3	4	5

Scaled [0, 1]

	Train			Test
	$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$
ϕ_1				
ϕ_2				
ϕ_3				

Standardized $\mathcal{N}(0, 1)$

	Train			Test
	$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$
ϕ_1				
ϕ_2				
ϕ_3				

Complete the two tables as explained below. Show your work.

1. Scale the features in the dataset to be between [0, 1]. Show the resulting scaled dataset in the second table.
2. Standardize the features in the dataset. Show the resulting standardized dataset in the third table. For the standardized values, you do not have to compute the final numbers, you can leave them in fractional form.

3 Vectorization (∗) (10 points)

Let $\phi(\mathbf{x}_1), \phi(\mathbf{x}_2), ..., \phi(\mathbf{x}_N)$ be the feature vectors of N training examples, represented as column vectors, i.e. $\phi(\mathbf{x}_n) \in R^{K\times 1}$, and let $\mathbf{t} = [t_1, t_2, ..., t_N]^T$ be the column vector of their

[1] https://en.wikipedia.org/wiki/Poisson_distribution

labels. To obtain the MLE estimation $\mathbf{w}_{ML}$ shown on slide 61 in class, we assumed in the hand notes that:

$$\sum_{n=1}^{N} \phi(\mathbf{x}_n)\phi(\mathbf{x}_n)^T = \Phi^T\Phi \tag{2}$$

$$\sum_{n=1}^{N} t_n\phi(\mathbf{x}_n)^T = \mathbf{t}^T\Phi \tag{3}$$

where Φ is the data matrix (also called design matrix) that contains the feature vectors on its rows, i.e. row number n in Φ is equal with $\phi(\mathbf{x}_n)^T$. Prove the identities at 2 and 3.

4 Submission

Submit your responses on Canvas as one file named `theory.pdf`. It is recommended to use an editor such as Latex or Word or Jupyter-Notebook that allows editing and proper formatting of equations. Alternatively, if you choose to write your solutions on paper, submit an electronic scan / photo of it on Canvas. Make sure that your writing is legible and easy to read.