CS 760: Machine Learning - Fall 2020
Homework 1: Review
Due : 09/24/2020

Zijie Zhang

September 22, 2020

Problem 1

Proof. By the definition, $\mathbb{R}^D$ is a subspace if for every a, $b \in \mathbb{R}$ and every $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^D$, $a\boldsymbol{u} + b\boldsymbol{v} \in \mathbb{R}^D$.
It is trivial. $\qquad \square$

Problem 2

Proof.

(a) $\boldsymbol{x} = (-1, \cdots, -1) \in \mathbb{R}^D$. The element-wise square roots is $(i, \cdots, i) \notin \mathbb{R}^D$.
$\mathbb{R}^D$ is not closed under element-wise square roots.

(b) $\mathbb{C}^D$.

$\qquad \square$

Problem 3

Proof. For every element $\boldsymbol{x}$, $\boldsymbol{y} \in \mathbb{U}$. They can be represented by a linear combination of $\boldsymbol{u}_1, \cdots, \boldsymbol{u}_R$.

$$\boldsymbol{x} = \alpha_1 \boldsymbol{u}_1 + \cdots + \alpha_R \boldsymbol{u}_R$$

$$\boldsymbol{y} = \beta_1 \boldsymbol{u}_1 + \cdots + \beta_R \boldsymbol{u}_R$$

For every a, $b \in \mathbb{R}$,

$$a\boldsymbol{x} + b\boldsymbol{y} = (a\alpha_1 + b\beta_1)\boldsymbol{u}_1 + \cdots + (a\alpha_R + b\beta_R)\boldsymbol{u}_R$$

$a\boldsymbol{x} + b\boldsymbol{y} \in \mathbb{U}$, so $\mathbb{U}$ is a subspace. $\qquad \square$

Problem 4

Proof.

(a) By the *Bayes rule*,

$$\mathbb{P}(\text{Have diabetes}|\text{These genes inactive})$$
$$= \frac{\mathbb{P}(\text{These genes inactive}|\text{Have diabetes}) \cdot \mathbb{P}(\text{Have diabetes})}{\mathbb{P}(\text{These genes inactive})}$$

(b) I need to know $\mathbb{P}(\text{These genes inactive})$.

(c) If the probability that these three genes inactive is very small, I should be concerned.

$\qquad \square$

Problem 5

Proof.

$$\mathbb{P}(x|\theta) = \begin{cases} \theta e^{-\theta(x-t_0)} & , x \geqslant t_0 \\ 0 & , x \leqslant t_0 \end{cases}$$

- t_0 *is the minimal time delay*: $\mathbb{P}(x \leqslant t_0|\theta) = 0$

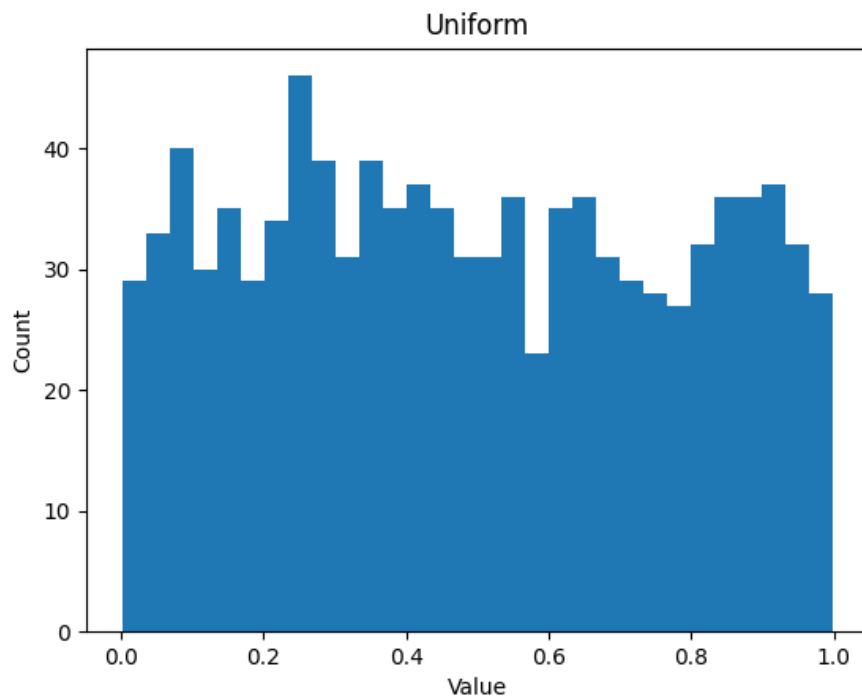- *Larger delays are rarer than shorter ones*: For all $x, y \in [t_0, \infty]$, $x < y$,

$$\frac{\mathbb{P}(y|\theta)}{\mathbb{P}(x|\theta)} = \frac{\theta e^{-\theta(x-t_0)}}{\theta e^{-\theta(y-t_0)}}$$
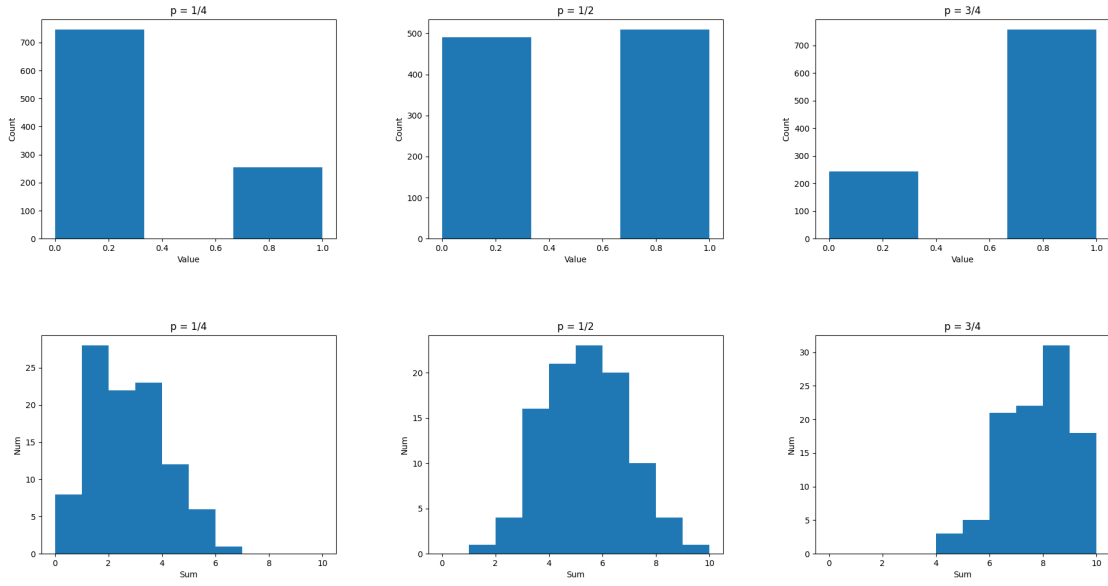$$= e^{-\theta(x-y)}$$
$$< 1$$

-

□

Problem 6

Proof.

(a) It doesn't look fairly uniform.

(b) $y_i \sim \text{Bernoulli(p)}$

(c) Yes.

(d) $z_k \sim \text{B(n,p)}$

(e) Yes.

□

p = 1/4

p = 1/2

p = 3/4

p = 1/4

p = 1/2

p = 3/4

Problem 7

Proof.

(a)
$$\nabla \ell(\boldsymbol{\theta}) = \sum_{i=1}^{N} \boldsymbol{x}_i \left(y_i - \frac{1}{1 + e^{-\boldsymbol{\theta}^T \boldsymbol{x}_i}} \right)$$

(b)
$$\boldsymbol{H} = \nabla^2 \ell(\boldsymbol{\theta}) = -\sum_{i=1}^{N} \frac{e^{\boldsymbol{\theta}^T \boldsymbol{x}_i}}{\left(1 + e^{\boldsymbol{\theta}^T \boldsymbol{x}_i}\right)^2} \cdot \begin{pmatrix} \boldsymbol{x}_{i_1}^2 & \boldsymbol{x}_{i_1} \boldsymbol{x}_{i_2} & \cdots & \boldsymbol{x}_{i_1} \boldsymbol{x}_{i_n} \\ \vdots & & & \vdots \\ \boldsymbol{x}_{i_n} \boldsymbol{x}_{i_1} & \boldsymbol{x}_{i_n} \boldsymbol{x}_{i_2} & \cdots & \boldsymbol{x}_{i_n}^2 \end{pmatrix}$$

(c) $\ell(\boldsymbol{\theta})$ is a scalar.
Its gradient $\nabla \ell(\boldsymbol{\theta})$ is a vector.
Its Hessian $\nabla^2 \ell(\boldsymbol{\theta})$ is a matrix.

□