Departamento de Ingeniería Matemática
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
UNIVERSIDAD DE CHILE

Curso: Aprendizaje de Máquinas, MA5204
Profesor: Felipe Tobar
Profesores Auxiliares: V. Faraggi, F. Fêtis, B. Moreno, F. Vásquez, A. Wortsman
Fecha de publicación: 09/06/21

Tarea Teórica #3

Dedicación recomendada: 10 horas

Fecha de entrega: 25 de Junio

Instrucciones: La tarea es individual. El formato de entrega es libre (hoja escaneada, Latex, hoja digital, etc), solo se pide que sea legible. Sea riguroso con sus demostraciones. Dudas y consultas por el foro de Ucursos.
Segunda Versión Cambios: En la P2. b) específica las dimensiones deben cambiar y se agrega un hint.

P1. Marco Teórico: Dadas n observaciones $i.i.d$ $\{x_i, y_i\}_{i=1}^n$, con $x_i, y_i \in \mathbb{R}$, del modelo:

$$y_i = r(x_i) + \epsilon_i, \quad i \in \{1, .., n\},$$

buscamos estimar la función r con una función $\hat{r}$. Un primer enfoque para resolver el problema anterior, sería considerar un estimador $\hat{r}$ de la forma:

$$\hat{r}(x) = \sum_{i=1}^{n} w(x, x_i) y_i, \tag{1}$$

donde $w(\cdot) : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ es una función base. Distintas elecciones de w entregan distintos estimadores. En esta pregunta nos concentraremos en el estudio de funciones base construidas con kernels.

Dado un kernel $K : \mathbb{R} \to \mathbb{R}$, con $\int K(u)du = 1$ y $K(x) = K(-x)$, y un ancho de banda h, definimos la función base:

$$w(x, x_i) = \frac{K\left(\frac{x - x_i}{h}\right)}{\sum_{j=1}^{n} K\left(\frac{x_j - x}{h}\right)}$$

a) Reemplace esta función base en (1). ¿Cómo interpreta este estimador para un Kernel cualquiera? ¿Cómo lo interpretaría si se eligiera el Kernel Gaussiano?

b) Se definen los estimadores:

$$\hat{p}(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h} K\left(\frac{x - x_i}{h}\right),$$

y:

$$\hat{p}(x, y) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h^2} K\left(\frac{x - x_i}{h}\right) K\left(\frac{y - y_i}{h}\right)$$

Por último, se define:

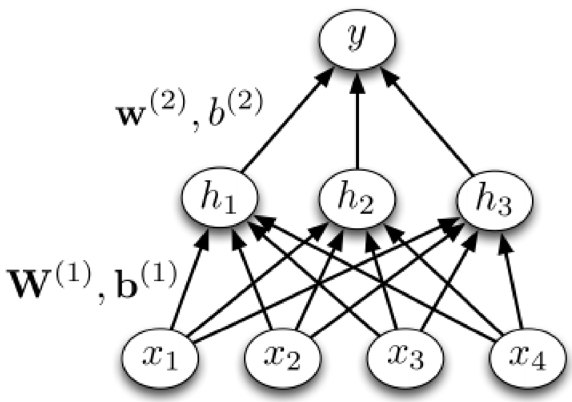
$$\hat{f}(x) = \frac{\int y \hat{p}(x, y) dy}{\hat{p}(x)}.$$

Suponiendo que $\displaystyle\int uK(u)du = 0$, demuestre que $\hat{f}$ corresponde al estimador encontrado en a).

P2. Redes Neuronales: En este problema, se busca diseñar dos redes neuronal a mano para que cada una cumpla con ciertos objetivos. Cuándo nos referimos por *diseñar a mano*, queremos decir que deberán elegir cada uno los pesos de las matrices $\mathbf{W}^{(k)}$ y $\mathbf{b}^{(k)}$ o explicitar un conjunto de condiciones necesarias para elegirlos.

La primera red que se quiere diseñar tiene que poder verificar si un elemento esta contenido en una lista de tamaño $n-1$. Para esto, se entrega un vector $\mathbf{x} \in \mathbb{R}^n$ donde los primeros $n-1$ elementos corresponden a la *lista* y un n-ésimo elemento, donde $x_i \in \mathbb{Z}$. Asuma que todos los elementos son distintos.

a) Primero, diseñe la red neuronal para el caso $n = 4$ siguiendo la siguiente arquitectura:



Utilice solamente la siguiente función de activación:

$$\phi(z) = \mathbb{I}(z \in [-1,1]) = \begin{cases} 1 & \text{si } z \in [-1,1] \\ 0 & \text{si } z \notin [-1,1] \end{cases}$$

b) Ahora, extienda para el resultado anterior para el caso de un n cualquiera. Es necesario modificar las dimensiones de $\mathbf{x}, \mathbf{W^{(1)}}, \mathbf{b^{(1)}}, \mathbf{w^{(2)}}, b^{(2)}$ de manera acorde. El output y sigue siendo escalar. Hint: Piense primero en cambiar las dos dimensiones de $\mathbf{W^{(1)}}$.

La segunda red que se quiere diseñar tiene que poder verificar si una lista es una permutación de otra. Al igual que para la red anterior, la entrada corresponde a un vector $\mathbf{x} \in \mathbb{R}^n$, donde n es siempre par, los primeros $\frac{n}{2}$ elementos corresponden a una lista, $l_1 = [x_1, x_2, ...x_{\frac{n}{2}}]$ y el resto a otra $l_2 = [x_{\frac{n}{2}}, x_{\frac{n}{2}+1}, ...x_n]$. Entonces, se busca determinar si l_2 es una permutación de l_1.

Esta red debe estar compuesta de redes más pequeñas como la que verifica si un elemento esta contenido en una lista. Se pueden utilizar otras funciones de activación. No debe explicitar los pesos de toda la red pero cada red interna debe estar bien definida.

c) Primero, diseñe la red neuronal para el caso $n = 6$.

d) Ahora, extienda para el resultado anterior para el caso de un n cualquiera.