



UNIVERSIDAD
DE LA SERENA
CHILE

60

Primera prueba parcial

Asignatura: Inteligencia Artificial
25% de la nota total del curso

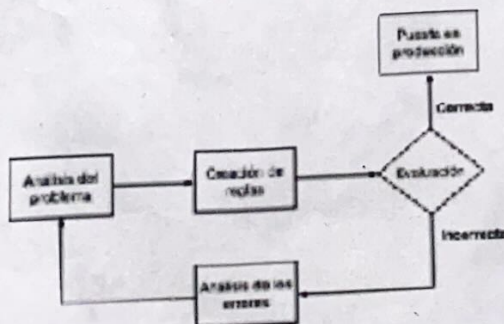
Nombre: SAVANDO ANTONIO
Rut: 19.505.884-9

Apellidos: PALMA PIZARRO

Instrucciones:

- Use cualquier lápiz de tinta.
- Responda las preguntas en las hojas adicionales, indicando claramente el numero de la pregunta.
- En cada hoja de respuesta agregue su nombre y rut.

- **Pregunta 1 [5 puntos]** : Indique las categorías y defina las subcategorías en las que se pueden clasificar los modelos de machine learning.
- **Pregunta 2 [10 puntos]** : Responda brevemente las siguientes preguntas:
 - ¿En qué situación puede no ser recomendable utilizar técnicas de Machine Learning?
 - ¿En qué categorías fundamentales se pueden clasificar los algoritmos basados en aprendizaje supervisado?
 - Los sistemas basados en aprendizaje batch funcionan bien para...
 - En los algoritmos de aprendizaje basado en modelos ¿Cuál es el rol de los parámetros?
- **Pregunta 3 [10 puntos]** : Represente gráficamente el porque no es posible usar la función de costo de la regresión lineal para entrenar a la regresión logística.
- **Pregunta 4 [5 puntos]** : Qué elementos debería cambiarse del siguiente proceso. Para que fuera un filtro de spam basado en machine learning.



○ **Pregunta 5 [10 puntos]** : Explique el proceso de optimización del método del gradiente descendente.

○ **Pregunta 6 [20 puntos]** : Describa paso a paso el proceso de entrenamiento de la **Regresión Lineal**.

Hipotesis

$$h_{\theta}(x) = \theta_0 + \theta_1 x$$

Parametros

$$\theta_0, \theta_1$$

Función de costo

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2$$

Cuando prima

1) los modelos de machine learning ordenamos en tres categorías como

- supervisado
- no supervisado
- semi supervisado.

no hay que olvidar

--- ~~FACT~~

Hay tres subcategorías donde se clasifican.

- Aprendizaje batch: que usa de utilizar todo el conjunto de datos, no incremental.
- Aprendizaje online: que se construye por ser un modelo incremental, se aprende cada vez que llegan datos.

7 por último lo que se bien según categorizar.

4/5

- aprendizaje basado en decisión: que crean una medida de similitud.
- aprendizaje basado en modelos: que crean ajustar sus parámetros para que el modelo converja.

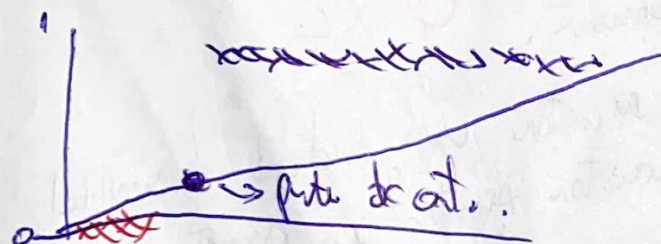
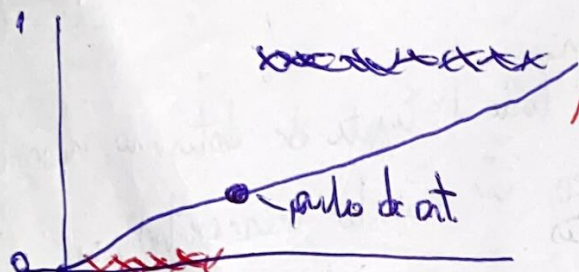
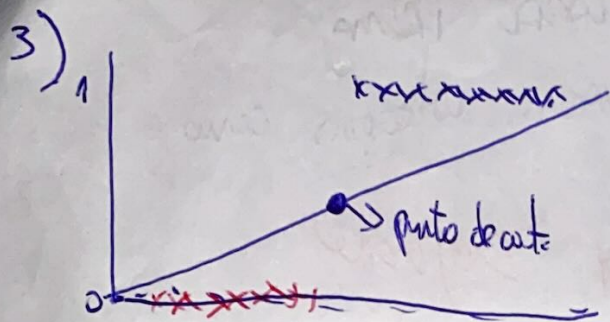
2) - no es recomendable usar machine learning cuando es un modelo existen reglas o flujos establecidos.

- pueden ser de 2 tipos de regresión o clasificación.

- son utilizados cuando se necesita crear todo el dataset que se tienen, además de ser rápidos y útiles, en otros palabras cuando se necesita un modelo rápido de hacer.

10/10

- los parámetros sirven para converger el modelo de otras formas y ven más resultados, un ejemplo es la tasa de error de algunos modelos o de aprendizaje.



El error consiste en que modelos de regresión lineal se basa en pendiente para ~~clasificar~~ en la tendencia de los datos, en un modelo de clasificación es difícil clasificar por el punto medio de clasificación, si el conjunto de datos es mayor para un tipo de label va a tender el modelo a ser label conponiendo con el contrario, esto dificulta en gran medida la clasificación

Paralelo

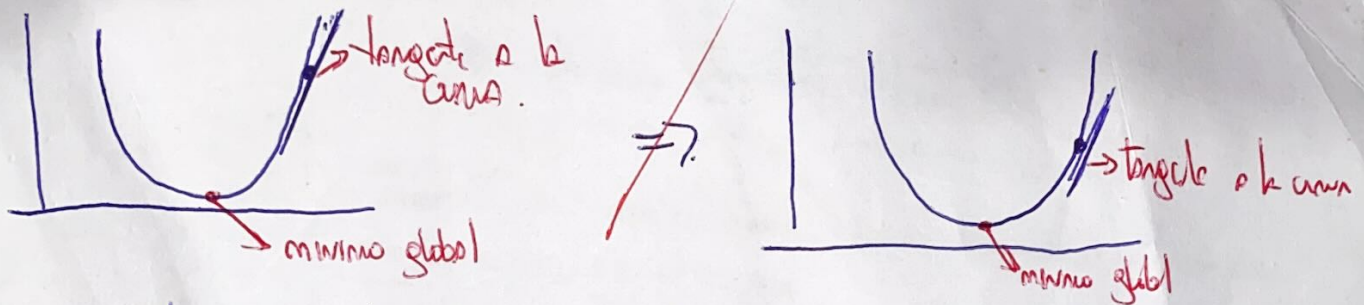
10/10

4) Se debe combinar específicamente la creación de reglas por un algoritmo de machine learning (un entrenamiento) etc. en la práctica de enumerar todos los reglas posibles en vez de hacerlas a mano, donde se optimiza el proceso de enumerar las reglas de un modelo.

FAL Diagrama

4/5

- 5) el método de optimización del gradiente descendente es el que se encarga de minimizar el error de modelo, donde se dibujan diagramas parecidos para disminuir el error.



en cada paso de optimización se debe ir disminuyendo el valor de la curva para hallar el mínimo global de la función de costo. para el ejemplo el punto más mínimo sería el mínimo por ~~negro~~ rojo. en ejemplo sería si la parte mínima global es el punto 0, la tangente de la curva sería 0 dando un error nulo.

- 6) 1) se asignan aleatoriamente valores a los parámetros θ_0 y θ_1 de la función hipotética, donde x corresponde a valores a entrenar para hallar el label correspondiente, a toda la conjunto de datos.
 2) luego se ingresa a la función de costo todos los datos con el valor predicho. esto para ver el error del modelo a producir.
 3) el resultado de 2 se usa como optimización del error correspondiente del modelo e ir modificando los parámetros correspondientes cambiando el descenso del gradiente... e ir ~~minimizando~~ minimizando el error a su mínimo global.
 4) según los pasos anteriores hasta que el modelo converge según la cantidad de los datos.



UNIVERSIDAD
DE LA SERENA
CHILE

49

Primera prueba parcial

Asignatura: Inteligencia Artificial
25% de la nota total del curso

Nombre: ~~Eduardo~~ ~~Pamela~~ ~~Pamela~~ Apellidos: ~~Palma~~ ~~Pamela~~
Rut: 19.505.884-9

Instrucciones:

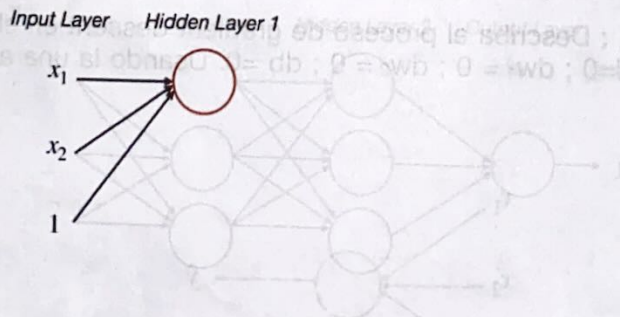
- Use cualquier lápiz de tinta.
- Responda las preguntas en las hojas adicionales, indicando claramente el numero de la pregunta.
- En cada hoja de respuesta agregue su nombre y rut.

○ **Pregunta 1 [10 puntos]** : Explicar porque Deep Learning se considera End-to-End Learning.

○ **Pregunta 2 [10 puntos]** : Responda brevemente las siguientes preguntas:

- ¿Porque resurgen las RNA ?
- ¿Indique las limitaciones de la neurona M-P?
- ¿Cuál es la diferencia fundamental entre el perceptron y MLP?
- ¿Que es la TLU?

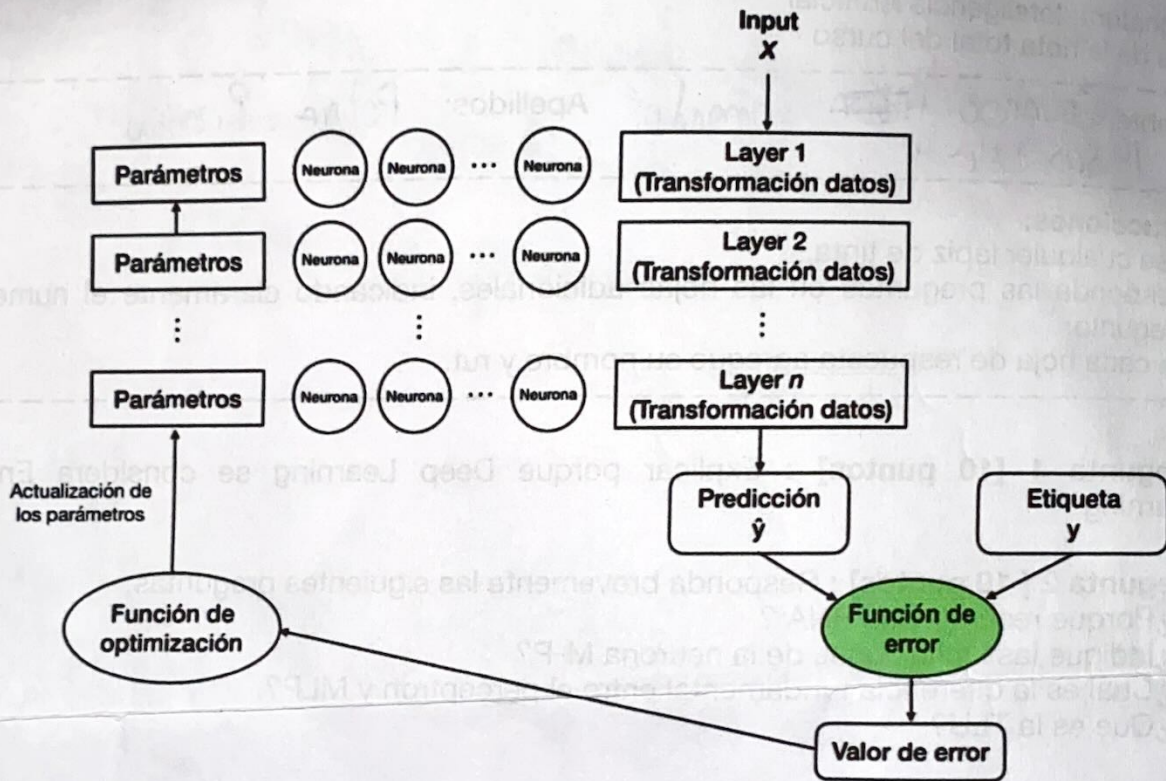
○ **Pregunta 3 [10 puntos]** : Describa paso a paso el proceso de Forward Propagation, usando la siguiente red:



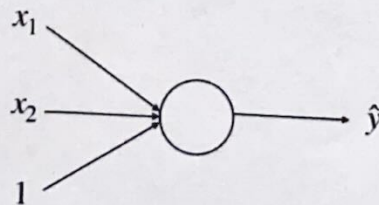
○ **Pregunta 4 [10 puntos]** : Explique la construcción de la función de costo (crossentropy) del MLP. Teniendo $y=1$ como la clase positiva e $y=0$ la clase negativa. Apoyase en los gráficos.

$$\text{Función de coste: } J(W, B) = -\frac{1}{m} \sum_{i=1}^m y^{(i)} \ln(\hat{y}^{(i)}) + (1 - y^{(i)}) \ln(1 - \hat{y}^{(i)})$$

o **Pregunta 5 [10 puntos]** : Explique el siguiente diagrama, en detalle.



o **Pregunta 6 [10 puntos]** : Describa el proceso de gradient descent en el MLP. Partiendo de la siguientes variables: $J=0$; $dw_1 = 0$; $dw_2 = 0$; $db = 0$. Usando la una arquitectura siempre de 1 TLU:



Pregunta 1.

se determina end-to-end learning debido a que los modelos de deep learning todo las características se hacen de un input para por el algoritmo de deep learning y emerge una salida. esto se hace en ciclo para que el modelo aprende de fin a fin. a dos problemas aprende de producir o producción.

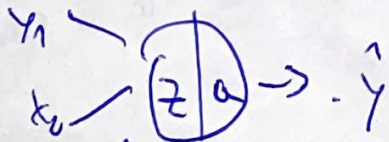
Pregunta 2.

- los RNA surgen por varias razones entre la mas destacadas se encuentran.
 - mayor acceso a los datos, hoy en dia existen grandes cantidades de datos.
 - Poder computacional mas barato gracias a las GPU.
 - los modelos de RNA ahora son mas complejos ademas de mejores.

- los limitaciones notran en su base que no resuelve problemas complejos ademas que solo produce valores 1, 0 sin un rango como ademas de no contienen un Bias.

- el perceptron o solo una neurona +1U, permite clasificar pero para problemas tipo XOR, no resulta conveniente, ademas de su limitacion como funcion de activacion y la simplicidad de la neurona. en cambio la MLP contiene varias neuronas, 2LU funcion de activacion mas conveniente y algoritmos de propagacion.

- En una neurona que contiene un z que es la suma ponderada de los pesos y por lo tanto, como ademas la funcion de activacion.



fragments 3.

- se ingresa todos los input a la red. para cada neurona, es la primera capa
- para cada neurona se determina una suma ponderada y se ingresa en su función de activación para la primera capa. Hacer lo mismo para las siguientes capas.

$$z = x_1 w_1 + x_2 w_2 + y.$$

$$a = a(z)$$

$$z_2 = w_1 w_1 + w_2 w_2 + \dots + w_n w_n + \text{Bias}$$

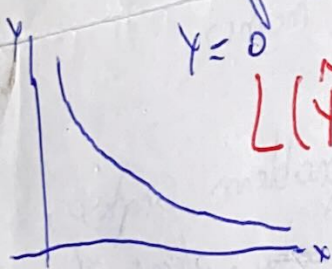
$$a = \sigma(z_2)$$

$$Q = Q(t_2)$$

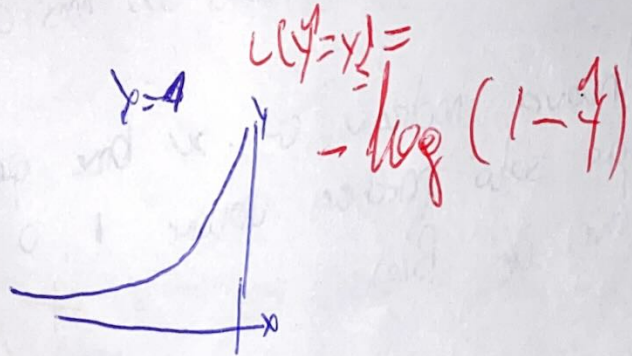
- Esto se repite para cada neurona en cada capa.
- El resultado de a se ingresa a la siguiente capa. Se repite las veces correspondientes sino podemos finiquitar de neurona.
- ~~Se repite~~ se repite el proceso en cada capa hasta la capa de salida.
- se repite la suma ponderada, función de activación, pero determinamos la producción en solo la capa de salida que costo de la neurona.

pregunta 4.

b. Finnen de beste perimete tussen 0 van 4 en van de conus se spino
b. met en paarden.



$$L(\hat{y}, y) = -\log(\hat{y})$$



$$L(y=x) = -\log(1-q)$$

Si el nodo se equivoca en la predicción de la clase correspondiente, la función de costo determina el grado de la confusión. Si el valor de error es 0, se determina una función de costo para un nodo particular. Si el valor es 0 o no. Si el modelo predice 0 y es 1 se con el error de 1 nodo en valor numérico.

Pregunta 5.

El diagrama representa un MLP con un input asociado. La red consta de 4 capas. El proceso es el siguiente.

- se ingresa a la red todos los valores de input
- se realizan las transformaciones correspondientes (como una ponderación, como la función de activación) para cada capa en la red. \rightarrow ¿el error?
- de como resultó una predicción
 - se ingresa la predicción como el bñl real a la función de error.
 - se determina un error para saber qué le sucede a la red
- se utiliza su función de optimización para aprender y determinar los parámetros adecuados para mejorar la predicción.
- se actualizan los pesos de los parámetros y entra en bucle nuevamente.

Pregunta 6

El gradiente descendente es la función de optimización para mejorar y obtener un mejor aprendizaje de la red, para el ejemplo se creó con una muestra 1×1 y 3 parámetros de entrada, el proceso que sigue es el siguiente.

- para cada dato de entrenamiento se ingresa y se realiza el proceso correspondiente con el error obtenido.
- se aplica la derivada para los 3 parámetros w_1, w_2 y b , donde γ corresponde al error acumulado de la red, esto es una suma de los errores obtenidos para cada dato y se suma el error ponderado acumulado para cada predicción.
- realizando los cálculos parciales se actualizan los parámetros de la red. \rightarrow ~~se actualizan los parámetros de la red~~