



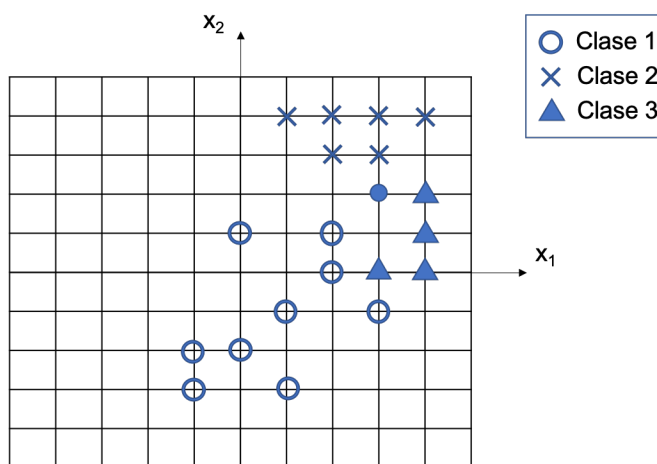
I302 - Aprendizaje Automático y Aprendizaje Profundo

1^{er} Semestre 2024

Modelo Parcial 1

-
1. (1.5 puntos) Sea un proceso estocástico que genera muestras (x, y) tal que $y = \hat{y}(x, w) + e(x)$, donde $\hat{y}(x, w)$ es el producto interno entre un vector de features $\phi(x)$ y un vector de parámetros w , y $e(x)$ es una fuente de ruido Gaussiano con distribución $\mathcal{N}(0, \sigma(x)^2)$ (es decir, el desvío estándar del ruido Gaussiano es función de la variable aleatoria x). Si tenemos un conjunto de muestras de entrenamiento con N muestras $\mathcal{D} = \{(x_i, y_i)\}$, con $i = 1, \dots, N$, derivar una expresión analítica para el estimador de máxima verosimilitud del vector de parámetros w , con respecto al conjunto de muestras \mathcal{D} .
 2. (2 puntos) Demostrar que regresión logística binaria con regularización L2 (con $\lambda > 0$) tiene un mínimo global que es único.
 3. (0.5 puntos) ¿Cuánto será el error de entrenamiento de un árbol de regresión (árbol de decisión aplicado a un problema de regresión), con restricción aleatoria de los features para determinar las particiones (node splitting), si no se limita la profundidad del árbol? Justifique su respuesta.
 4. (1 punto) Marque cuáles de las siguientes opciones son correctas. En regresión lineal, si agrego un feature nuevo:
 - (a) El error de test se reducirá o se mantendrá constante.
 - (b) El error de entrenamiento se reducirá o se mantendrá constante.
 - (c) El error de validación puede bajar.
 - (d) El error de validación puede aumentar.

5. (1 punto) Durante el algoritmo backpropagation, a medida que el gradiente fluye hacia atrás a través de una función de activación sigmoide, el gradiente siempre:
- Aumenta en magnitud y mantiene el signo.
 - Aumenta en magnitud e invierte el signo.
 - Disminuye en magnitud y mantiene el signo.
 - Disminuye en magnitud e invierte el signo.
6. (1 punto) Para las siguientes afirmaciones sobre árboles de decisión, indique si son verdaderas o falsas.
- Suelen mantener una gran estabilidad al agregar más datos.
 - Son aproximadores universales.
 - Son robustos a outliers.
 - La técnica de pruning en general aumenta el sobreajuste del modelo.
7. (1.5 puntos) Demostrar que un modelo de clasificación Linear Discriminant Analysis (LDA) resulta en fronteras de clasificación entre una clase y otra lineales en el espacio de features.
8. (0.5 punto) Sea un problema de clasificación multiclase, con dos features (x_1 y x_2) y un conjunto de entrenamiento como el que se muestra abajo. Calcular las probabilidades posteriores de cada clase para el punto marcado con el círculo relleno, si se aplicara el algoritmo KNN con $K = 7$.



9. (0.5 puntos) Marque cuales de las siguientes afirmaciones son correctas. ¿Por qué es importante inicializar los pesos de una red neuronal densamente conectada con valores aleatorios cercanos a cero?
- Porque queremos que la red este regularizada.
 - Porque la aleatoriedad mejora el set de datos.

- (c) Porque queremos que muchas neuronas esten activas al inicio del entrenamiento.
 - (d) Porque no queremos que los gradientes sean demasiado grandes al inicio del entrenamiento.
10. (0.5 puntos) Marque cuales de las siguientes afirmaciones son correctas. En regresion lineal, al normalizar los features con la media y la varianza de cada feature:
- (a) Los features normalizados tendrán valores entre 0 y 1.
 - (b) Al evaluar el error predictivo del modelo entrenado sobre un conjunto de validación, hay que recalculr los factores de normalización antes de hacer las predicciones.
 - (c) Esto mejorará el numero de condición de la pseudo-inversa de la matriz de diseño.
 - (d) La normalización se hace para que los datos sean fáciles de graficar.