I302 - Aprendizaje Automático y Aprendizaje Profundo

2^{do} Semestre 2024

Guía de Ejercicios 1

---

1. Sea una matriz A $\in \mathbb{R}^{m \times n}$ de rango completo, se define el número de condición de la matriz A como:
$$\kappa(A) \triangleq ||A||.||A^\dagger||$$
Demostrar que:
$$\kappa(A) = \frac{\sigma_{max}(A)}{\sigma_{min}(A)}$$
Sugerencia: primero demostrarlo para el caso de $m = n$ (matriz cuadrada), y luego para los casos $m > n$ y $m < n$. Recordar las definiciones de $A^\dagger$ para cada caso y usar la descomposicón en valores singulares de A.

2. Sea $f(x) : \mathbb{R}^n \to \mathbb{R}$ una función continua y doblemente diferénciable y $g(z) : \mathbb{R} \to \mathbb{R}$ una función continua, doblemente diferenciable y monotónicamente creciente. Demostrar que si un punto x^* es mínimo local de $f(x)$, entonces también es un mínimo local de la función compuesta $h(x) = g(f(x))$.

   Sugerencia: plantear las condiciones de optimalidad local de primer y segundo orden para $f(x)$ y $h(x)$, y verificar que si se cumplen para $f(x)$, entonces también se cumplen para $h(x)$. Si quiere, primero puede demostrarlo para el caso particular de $n = 1$ (es decir el caso en que x es un escalar) y luego generalizarlo para el caso en que x es un vector.

3. Derivar las ecuaciones para determinar los parámetros óptimos $\mathbf{w}^*$ de un modelo de Regresión Lineal aplicando el principio de máxima verosimilitud, asumiendo que las muestras son independientes e identicamente distribuidas con un ruido Gaussiano $\mathcal{N}(0, \sigma^2)$.

Demostrar que $\mathbf{w}^*$ es el óptimo global, que maximiza globalmente la verosimilitud (probabilidad conjunta) de los datos de entrenamiento.

4. Derivar las ecuaciones para determinar los parámetros óptimos $\mathbf{w}^*$ de un modelo de Regresión Lineal aplicando el principio de máxima verosimilitud, pero en este caso asumiendo además que la incertidumbre en el vector $\mathbf{w}$ está dado por la distribución de probabilidad:

$$P(\mathbf{w}) = \mathcal{N}(\mathbf{w}_0, s^2 I) \tag{1}$$

Donde $\mathbf{w}_0$ es un vector constante. Es decir, la distribución a-priori (antes de ver los datos) de $\mathbf{w}$ es una Gaussiana de media $\mathbf{w}_0$ y co-varianza $s^2 I$.

5. Derivar la función de costo "binary cross-entropy" para un problema de clasificación binaria con regularización L_2 sobre los parámetros w, con hiperparámetro λ, aplicando el principio de máxima verosimilitud sobre el conjunto de entrenamiento.