

Índice general

1. Introducción	5
1.1. Aprendizaje supervisado	5
1.1.1. Parámetros e hiperparámetros	6
1.1.2. Error de Bayes	7
1.2. Aprendizaje no supervisado	7
1.3. Aprendizaje por refuerzos o recompensas	7
2. Esperanza Condicional	8
2.1. Esperanza condicional respecto de una σ -álgebra	8
2.2. Esperanza condicional respecto de una variable	13
2.3. Esperanza condicional respecto de $X = x$	13
2.4. Distribución Condicional y Varianza Condicional	16
3. Regresión y Clasificación	18
3.1. Regresión	18
3.2. El teorema de Stone	18
3.3. Clasificación	21
3.3.1. Clasificación Binaria	21
3.3.2. Regla de Bayes	21
3.3.3. Reglas plug-in	23
3.3.4. Criterios de minimización del error	23
3.3.5. Reglas construidas a partir de una muestra	24
3.4. Clasificación por vecinos mas cercanos, k -NN	26
4. Clasificación lineal y particiones	27
4.1. Clasificación Lineal Univariada	27
4.2. Clasificación Lineal Multivariada	29
4.2.1. Discriminante lineal de Fisher	29
4.3. Reglas de clasificación basadas en particiones	30
4.3.1. Histogramas	32
5. Modelos lineales	33
5.1. Estimación 1	35
5.1.1. Estimación lineal insesgada de mínima varianza [ELIVM]	36
5.1.2. Conexión con mínimos cuadrados	37
5.1.3. Descomposición ortogonal del error. Estimación insesgada de la varianza	37
5.2. Modelos lineales con errores normales. Distribución de los estimadores	39
5.3. La prueba F	40
5.4. Regresión Logística	41
5.4.1. Ajuste de w por máxima verosimilitud	42
6. Teoría de Vapnik-Chervonenkis	44
6.1. Glivenko-Cantelli	45
6.2. Coeficiente de Fragmentación	48
6.2.1. Condición necesaria y suficiente	49
6.2.2. Elección de clasificadores	49
6.3. Aspectos combinatorios de la teoría de Vapnik-Chervonenkis	50
6.4. Error de resustitución	55

7. Redes Neuronales	57
7.1. El perceptrón de Rosenblatt	57
7.2. Redes con L capas	59
7.3. La función XOR y el sigmoide ReLu	60
7.3.1. Otras redes que no veremos	60
7.4. Particiones aleatorias con hiperplanos y redes neuronales	61
7.4.1. Consistencia	61
7.4.2. De particiones a redes neuronales de 2 capas	61
7.4.3. Consistencia de las redes con 1 capa	62
7.4.4. Dimensión de VC de las redes neuronales	63
8. Métodos de descenso por gradiente y gradiente estocástico	66
8.1. Descenso por gradiente, SGD	67
8.1.1. Algoritmo de Robbins-Monro	67
8.1.2. SGD con Momento	68
8.2. Cálculo de gradiente por propagación hacia atrás	68
8.2.1. Definiciones y Notaciones	68
8.2.2. Propagación hacia adelante	68
8.2.3. Propagación hacia atrás	69
8.2.4. Ejemplo Concreto	70
8.2.5. Actualización de Pesos y Sesgos	70
9. Algunos conceptos básicos de teoría de la medida	71
9.1. Teorema de cambio de variable	72
9.2. Integrales iteradas en \mathbb{R}^d	72
9.3. Clases monótonas y Teorema de Radon-Nikodym.	72
9.4. Desigualdad de Hoeffding	74