

Tabla de contenido

Prefacio	ix
1. Panorama de los sistemas de aprendizaje automático	1
Cuándo utilizar el aprendizaje automático	3
Casos de uso del aprendizaje automático	9
Cómo entender los sistemas de aprendizaje automático	12
El aprendizaje automático en la investigación y en la práctica	12
Los sistemas de aprendizaje automático y el software tradicional	22
Resumen	24
2. Introducción al diseño de sistemas de aprendizaje automático	25
Objetivos empresariales y de AA	26
Requisitos para los sistemas de AA	29
Confiabilidad	29
Escalabilidad	30
Capacidad de mantenimiento	31
Adaptabilidad	32
Proceso iterativo	32
Planteamiento de problemas de AA	34
Tipos de tareas del AA	35
Funciones objetivo	39
El intelecto frente a los datos	42
Resumen	45
3. Fundamentos de la ingeniería de datos	47
Las fuentes de datos	48
Formatos de datos	51
JSON	52
Los formatos de orden por filas frente a los de orden por columnas	52
Texto frente a formato binario	55
Modelos de datos	56
Modelo relacional	57
NoSQL	61
Datos estructurados frente a datos no estructurados	64
Motores de almacenamiento de datos y procesamiento	65
Procesamiento transaccional y analítico	66
ETL: Extraer, transformar y leer	69

Modos de flujo de datos	71
Datos que pasan por bases de datos	71
Datos que pasan a través de los servicios	71
Transferencia de datos a través del transporte de datos al instante	73
El procesamiento por lotes frente al procesamiento continuo	76
Resumen	78
4. Datos de formación.....	81
Muestreo	82
Muestreo no probabilístico	83
Muestreo aleatorio simple	84
Muestreo estratificado	84
Muestreo ponderado	85
Muestreo de depósito	86
Muestreo de importancia	87
Etiquetado	88
Etiquetas manuales	88
Etiquetas naturales	91
Cómo afrontar la falta de etiquetas	94
Desequilibrio de clases	103
Desafíos del desequilibrio de clases	104
Cómo afrontar el desequilibrio de clases	106
Aumento de datos	115
Transformaciones simples que conservan la etiqueta	115
Perturbación	116
Síntesis de datos	118
Resumen	120
5. Creación de atributos	121
Los atributos aprendidos frente a los diseñados	121
Operaciones comunes de creación de atributos	124
Manejo de valores faltantes	125
Escalado	127
Discretización	129
Codificación de atributos categóricos	130
Cruce de atributos	133
Incrustaciones posicionales discretas y continuas	134
Fuga de datos	136
Causas comunes de la fuga de datos	138
Detección de la fuga de datos	141
Creación de atributos buenos	142

Importancia de los atributos	142
Generalización de los atributos	145
Resumen	146
6. Desarrollo de modelos y evaluación fuera de línea.....	149
Desarrollo de modelos y formación	150
Evaluación de modelos de AA	150
Ensamblados	156
Rastreo de experimentos y control de versiones	162
Formación distribuida	167
AutoML	172
Evaluación del modelo fuera de línea	178
Referencias	179
Métodos de evaluación	181
Resumen	188
7. Servicio de predicción y puesta en marcha de modelos.....	191
Mitos sobre la instalación del aprendizaje automático	193
Mito 1: solo se instalan uno o dos modelos de AA a la vez	194
Mito 2: si no hacemos algo, el rendimiento del modelo seguirá igual	195
Mito 3: no necesitará actualizar mucho los modelos	195
Mito 4: la mayoría de los ingenieros de AA no necesitan preocuparse por la escala	196
La predicción por lotes frente a la predicción en línea	197
De la predicción por lotes a la predicción en línea	200
Unificación de la canalización por lotes y la canalización de flujo	203
Compresión del modelo	205
Factorización de bajo rango	206
Destilación del conocimiento	207
Poda	207
Cuantificación	208
AA en la nube y en el perímetro	211
Compilación y optimización de modelos para dispositivos periféricos	213
El AA en los navegadores	221
Resumen	222
8. Cambios y monitoreo de la distribución de los datos.....	225
Causas de los fallos en los sistemas de AA	226
Fallos del sistema informático	227
Fallos específicos del AA	228

Cambios en la distribución de los datos	236
Tipos de cambios en la distribución de los datos	237
Cambios generales en la distribución de los datos	241
Detección de cambios en la distribución de los datos	242
Cómo afrontar los cambios en la distribución de los datos	248
Monitoreo y observabilidad	250
Métricas específicas del AA	252
Herramientas de monitoreo	256
Observabilidad	260
Resumen	262
9. Aprendizaje continuo y pruebas en producción	265
Aprendizaje continuo	266
Reformación sin estado frente a formación con estado	267
El porqué del aprendizaje continuo	270
Desafíos del aprendizaje continuo	272
Las cuatro etapas del aprendizaje continuo	277
Frecuencia de actualización de los modelos	282
Pruebas en producción	285
Implementación en paralelo	286
Pruebas A/B	286
Versión canario	288
Experimentos intercalados	289
Bandidos	291
Resumen	295
10. Infraestructura y herramientas para MLOps.....	297
El almacenamiento y los cálculos	301
Nube pública frente a centros de datos privados	304
Entorno de desarrollo	306
Configuración del entorno de desarrollo	307
Estandarización de los entornos de desarrollo	310
Del desarrollo a la producción: contenedores	312
Gestión de recursos	315
Cron, planificadores y orquestadores	316
Gestión del flujo de trabajo de la ciencia de datos	319
La plataforma del AA	324
Puesta en marcha de los modelos	325
Almacén de modelos	326
Almacén de atributos	330
Crear frente a comprar	332

Resumen	335
11. El lado humano del aprendizaje automático.....	337
Experiencia del usuario	337
Garantizar la coherencia de la experiencia del usuario	338
Cómo evitar las predicciones «casi correctas»	338
Fallo suave	340
Estructura de equipo	340
Colaboración en equipos multifuncionales	341
Científicos de datos de principio a fin	341
IA responsable	345
IA irresponsable: caso de estudio	346
Un entorno para la IA responsable	353
Resumen	359
12. Epílogo.....	361