**Autores**

**JAMES – CAPITULO 2**

**HASTIE – CAPITULO 2**

**Aprendizaje Supervisado**

Entrada x, salida y

Entrenar modelo que prediga valor de a partir de x

Flujo de ML

1. Extracción (Ingesta, extracción, validación)
2. Procesamiento (ETL/ELT)
3. Entrenamiento (Sintonización, Función del Error)

* Optimizar función del error (encontrar pesos óptimos en los datos de entrenamiento)
* Evaluar el desempeño del modelo en los datos de validación
* Seleccionar el modelo con menor función de error en los datos de validación
* Evaluar modelo final en datos de prueba

1. Selección de Modelo
2. Evaluación
3. Interpretabilidad
4. Monitorización

Criterio del Error de Clasificador

Cálculo del Error Empírico de h en los datos de prueba

Cotas de Chernoff

Para clasificación, para un dato de prueba h comete un error con probabilidad

Entonces

Con probabilidad datos de prueba, se garantiza con probabilidad de por lo menos que

**Predicción**

es estimador de f

predicción para y

El accuracy de como predicción de y depende de 2 tipos de errores: Error reducible y Error irreducible

El Error Reducible se puede mejorar mediante apropiadas técnicas de aprendizaje supervisado

El Error Irreducible es la variabilidad asociada a que no se puede reducir

Dado estimador , features x

Se generan las predicciones

Error Reducible

Error Irreducible/varianza del error intrínseco

promedio/valor esperado de la diferencia al cuadrado entre el valor real y el valor predicho

Aprendizaje supervisado son técnicas para estimar para minimizar el Error Reducible

El Error Irreducible genera un límite superior del accuracy en las predicciones

**Inferencia**

Entender relación entre x features/predictores/variables dependientes

Entender cómo cambia en función de x

* Identificar features/predictores con mayor impacto con la variable dependiente
* Relación (correlación) entre feature con repuesta de y

**Métodos Paramétricos**

1. Asumir forma de
2. Entrenamiento de modelo (encontrar pesos de modelo
3. Estimar set de parámetros/pesos

Sobreajuste/Overfitting: modelos ajustados para seguir el error o ruido

**Métodos No Paramétricos**

No hacen suposiciones de la forma de

Buscan un estimador de lo más cercano a los datos

Como no se asume una forma de los modelos pueden ajustar bien diferentes formas de f

No se reduce el problema de estimar a obtener los estimadores de los pesos , entonces se necesitan muchos datos para generar estimados precisos de

**Trade-off entre Accuracy de la Predicción y la Interpretación**

Imagen que contiene Interfaz de usuario gráfica

Descripción generada automáticamente

Cuando el objetivo es la inferencia es mejor seleccionar modelos restrictivos y simples

**Aprendizaje Supervisado vs Aprendizaje No Supervisado**

Aprendizaje Supervisado: Entrenar(ajustar) pesos de un modelo que relaciona los features/predictores con la variable dependiente para predecir correctamente la respuesta de la variable dependiente para futuras observaciones no vistas (predicción) o para entender las relaciones entre la respuesta de la variable dependiente y los predictores(inferencia)

Aprendizaje No Supervisado: Entender relaciones entre los predictores/features/variables

Aprendizaje Semi Supervisado: Incorporar observaciones con respuesta de variable dependiente conocida y observaciones sin respuesta de variable dependiente en un modelo.

**Accuracy del Modelo**

**Métricas de Evaluación**

**MSE:** Error Cuadrático Medio

Sobreajuste/Overfitting: Cuando modelo buenos (error bajo), pero mal desempeño en los datos de prueba (error alto).

Modelo/Estimador de se entrena/ajusta al Error Intrínseco/Irreductible en los datos de entrenamiento y patrones que no existen en los datos de prueba, entonces no generaliza.

Las métricas de desempeño en datos de entrenamiento por lo general son ligeramente mejores que métricas de desempeño en los datos de prueba.

Técnicas de Sobremuestreo/ Resampling: Técnicas para estimar el mínimo global de la Función del Error

**Trade-off entre Sesgo/Varianza**

El promedio/valor esperado de la función del Error MSE para una observación en los datos de prueba se compone de 3 factores:

varianza de función/modelo evaluado en dato

sesgo al cuadrado de función/modelo evaluado en dato

varianza del Error Irreductible

El mejor modelo de ML es el que alcanza baja varianza, bajo sesgo de modelo/función

Varianza y Sesgo cuadrado de son siempre valores positivos, entonces MSE promedio/esperado no puede ser menor que la varianza del Error (Error Irreducible)

Varianza de : Cantidad por la que cambia el ajuste del modelo para diferentes datos de entrenamiento. El modelo estimado de perfecto no variaría entre diferentes datos de entrenamiento o muy poco.

Modelos con alta varianza son modelos que con pequeños cambios en los datos de entrenamiento resultan en cambios drásticos de los pesos del modelo estimado

Sesgo(cuadrado) d e: Error introducido por usar modelos simples/restrictivos con datos complejos/altamente no lineales

Modelos más flexibles/complejos tienen más varianza

Modelos más flexibles/complejos tienen menos sesgo

Gráfico

Descripción generada automáticamente

Técnicas de Resampleo (Validación Cruzada): Técnica para estimar Función del Error MSE usando datos de prueba

**Clasificación**

La Función del Error para cuantificar el accuracy del modelo en clasificación es el Error Rate

Error Rate: proporción de clasificaciones incorrectas.

clase predicha con función/modelo

variable indicadora

**Clasificador de Bayes**