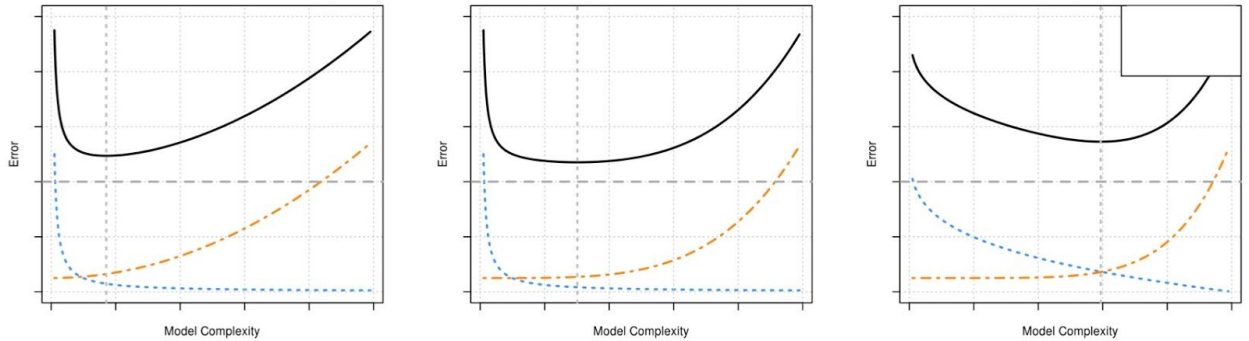


# Guía de ejercicios de Técnicas de Machine Learning para Análisis de Datos

## Junio 2020

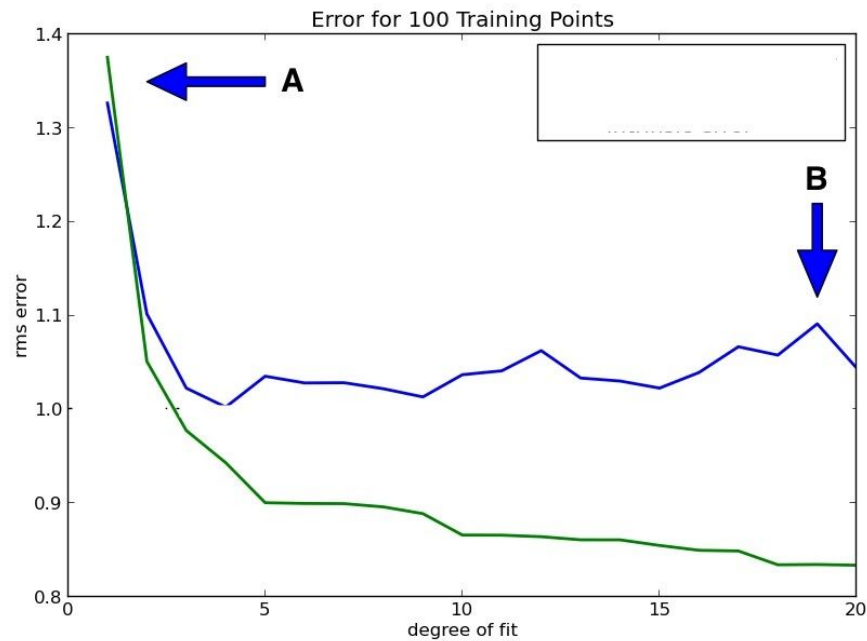
### Ejercicio 1

Explique las curvas graficadas en cada figura, incluyendo las líneas punteadas, y realice un análisis comparativo entre las figuras. Justifique el análisis.



### Ejercicio 2

1. Defina, explique funcionamiento y uso de la técnica de k-fold cross validation (CV).
2. Considere la siguiente gráfica:



- Identifique las curvas de CV y train.
- Para A y B indique si corresponden a alto sesgo y/o alta varianza. Justifique.

## Guía de ejercicios de Técnicas de Machine Learning para Análisis de Datos Junio 2020

### Ejercicio 3

Considere una matriz de confusión tal que positivo (P) corresponde a  $Y=1$  y negativo (N) a  $Y=0$ , siendo  $Y$  la variable a predecir. Defina la fórmula correspondiente a:

1. Error de clasificación
2. Tasa de precisión: proporción de positivos predichos correctamente clasificados.
3. Tasa de recuperación: proporción de positivos reales correctamente clasificados.

Suponga que la probabilidad  $P[Y=1|x]$  calculada es la siguiente (el color gris indica que  $Y=1$ ).

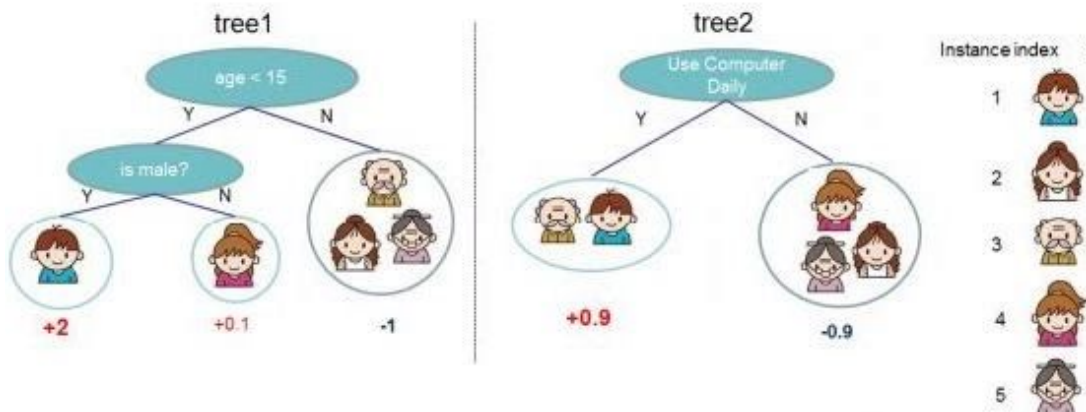
.98	.95	.93	.92	.90	.87	.85	.84	.81	.79	.77	.75	.73	.69	.65	.62	.58	.55
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

1. Si el umbral de clasificación aumenta, qué pasa con la precisión?
  - a. Tiende a aumentar
  - b. Tiende a disminuir
2. Si el umbral de clasificación aumenta, qué pasa con la recuperación?
  - a. Tiende a aumentar
  - b. Tiende a disminuir

Justifique su respuesta.

### Ejercicio 4

Explique el algoritmo de Boosting.



Considere el ejemplo de la figura. Asuma que la variable a predecir  $Y$  es +1 si el individuo es joven y -1 si es anciano. Cuál es la predicción para cada individuo? Cuál es el error?

**Guía de ejercicios de Técnicas de Machine Learning para Análisis de Datos**  
**Junio 2020**

**Ejercicio 5**

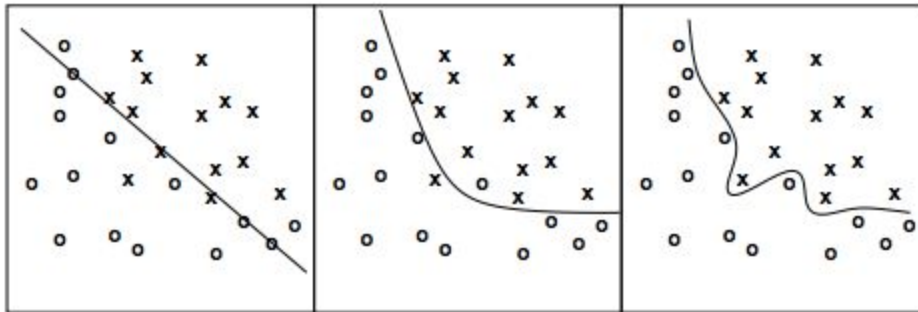
Explique cuál fue la hipótesis seleccionada en el obligatorio y el método que aplicó para obtenerla. Cómo le fue con esta hipótesis en la competencia? Cómo explica el resultado obtenido?

**Ejercicio 6**

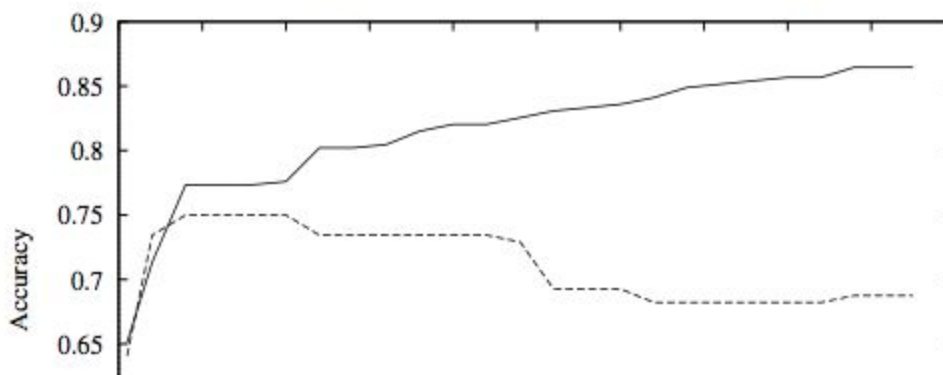
1. Explique la diferencia entre el enfoque generativo y el discriminativo.
2. Clasifique los siguientes algoritmos como generativo o discriminativo. Justifique.
  - a. Naïve Bayes
  - b. Regresión logística
  - c. Árboles de decisión

**Ejercicio 7**

1. Considere la siguiente figura.
  - a. Explique todo lo que se observa.
  - b. ¿Qué hipótesis seleccionaría? Justifique



2. La siguiente figura muestra dos curvas de exactitud o "accuracy".
  - a. Defina exactitud.
  - b. Identifique cada una de las curvas observadas.
  - c. ¿Qué hipótesis seleccionaría? Justifique.



**Guía de ejercicios de Técnicas de Machine Learning para Análisis de Datos**  
**Junio 2020**

**Ejercicio 3.**

1. Defina el concepto general de ensemble.
2. Considere la siguiente tabla.

Elemento	C1	C2	C3	C4	C5
A	*		*	*	*
B	*		*	*	*
C	*	*		*	*
D	*	*			
E		*	*		

La tabla muestra los resultados obtenidos por 5 clasificadores C1, C2, C3, C4 y C5, sobre un conjunto de datos compuesto por 5 elementos, identificados como A, B, C, D y E. Una \* indica que el elemento fue clasificado correctamente.

- a. Calcule el error cometido por el ensemble compuesto por los 5 clasificadores, asumiendo peso uniforme y la **esperanza del error** de los clasificadores.
- b. ¿Se verifica la propiedad fundamental del error del ensemble con respecto a los errores de los clasificadores? Justifique.
- c. En caso que la respuesta sea negativa, construya un ensemble a partir de los clasificadores dados que cumpla la propiedad. Justifique.

**Ejercicio 7**

1. Explique qué se entiende por regularización.
2. Explique Ridge (norma 2) y LASSO (norma 1).
3. En la siguiente tabla se muestran los coeficientes de una regresión sin (Coeficiente 1) y con (Coeficiente 2) regularización obtenida sobre un conjunto de datos. ¿De qué tipo de regularización se trata? Justifique.

Variable	Coeficiente 1	Coeficiente 2
(intercept)	0.08370835	0.07960114
duration	0.00000035	0.00000035
distance	0.00000004	0.00000004
fare	0.17503086	0.17800492
tolls	0.06266734	0.00000000
weekends	-0.02823731	0.00000000