

# Aprendizaje Automático

Depto. de Computación. FCEyN. UBA  
2º cuatrimestre de 2016

## Primer Parcial

El parcial se aprueba con 60 puntos o más.  
Poner nombre y apellido en todas las hojas.

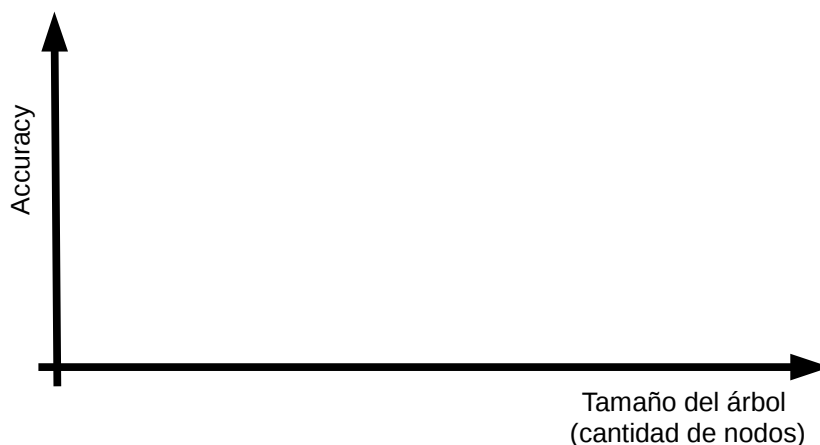
Nombre:				
Problema 1	Problema 2	Problema 3	Problema 4	Problema 5
Total:			Nota:	

**Problema 1. [40 puntos]** Definir brevemente exactamente 8 de los siguientes términos:

- Navaja de Occam
- Espacio de hipótesis
- Matriz de confusión
- Pruning (poda) en árboles de decisión
- Vectores de soporte (*support vectors*)
- Greedy forward selection
- Compromiso entre sesgo y varianza
- Stacking de clasificadores
- Principal component analysis (PCA)
- Compromiso entre *precision* y *recall*

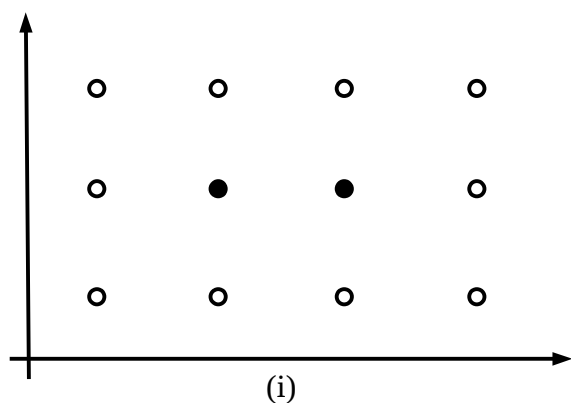
**Problema 2. [15 puntos]** Describir la técnica de clasificación conocida como Naive Bayes.

**Problema 3. [15 puntos]** Considerar el algoritmo de árboles de decisión sin poda. En la siguiente figura, dibujar dos curvas que describan aproximadamente cómo variaría la eficacia (*accuracy*) del modelo a medida que crece la cantidad de nodos, evaluada sobre 1) los datos de entrenamiento y 2) los datos de test, en ambos casos sin cross-validation. Justificar.

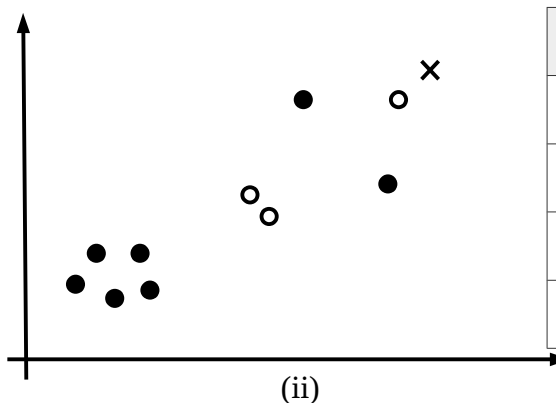


**Problema 4. [15 puntos]** Considerar el algoritmo de  $k$  vecinos más cercanos (kNN) para un conjunto de puntos bidimensionales, con dos clases posibles:  $\circ$  y  $\bullet$ . Tener en cuenta la distancia euclídeana.

- En la figura (i), dibujar el Diagrama de Voronoi correspondiente a  $k=1$ .
- Teniendo en cuenta los datos de la figura (ii), estimar la probabilidad de pertenencia a la clase  $\circ$  de la nueva instancia marcada con "X", cuando  $k=1$ ,  $k=3$ ,  $k=5$  y  $k=N$ , donde  $N$  es el tamaño del conjunto de datos de entrenamiento. Ingresar los resultados en la tabla de la derecha.



(i)



(ii)

$k$	$P(c(X)=\circ)$
1	
3	
5	
$N$	

**Problema 5. [15 puntos]** ¿Por qué, en Random Forest, en cada nodo de cada árbol se considera sólo un subconjunto de atributos elegidos al azar?