

Cuestionario 1 AA

1. Suponga que disponemos de un conjunto de 1000 muestras etiquetadas de un problema de clasificación binaria a partir de las cuales queremos entrenar 5 modelos distintos y elegir el mejor de ellos. Cada modelo es entrenado usando un algoritmo específico y una clase de funciones finita de tamaño M . Describa qué decisiones tomaría para ello y calcule la mejor cota para el error E_{out} del modelo elegido.

2. Suponga que le encargan escribir un programa para etiquetar muestras sin etiqueta en un problema de clasificación binario. No se conoce la verdadera función de etiquetado f , pero si se conoce la distribución de las muestras P . ¿Cómo lo haría para garantizar un error de clasificación lo más bajo posible?

3. Considere la función de error definida para una muestra (x,y) por $E(w) = (\max(0, 1 - yw^T x))^2$. Deducir la regla de adaptación de gradiente descendente para el parámetro w . Usar el resultado para deducir la regla de adaptación para la función de error $1/N \sum_{n=1}^N E_n(w)$ asociada al promedio de n muestras.

4. Cuando resolvemos un problema de aprendizaje desde datos, diga cuáles de las siguientes opciones son correctas:

1.- Después de aprender obtenemos una función g , con garantía de que aproxima bien a la función f , fuera de la muestra.

2.- Después de aprender obtenemos una función g , que con alta probabilidad dicha aproximará a la función f , bien fuera de la muestra.

3.- Logramos una de estas dos cosas:

a) {Obtendremos una función g que con alta probabilidad aproximará a la función f bien fuera de la muestra.

b) Diremos que hemos fallado

Justificar la elección.

5. Suponga el siguiente escenario. Un matemático que no cree en la teoría del aprendizaje desde datos le reta a que averigüe una función de etiquetado $f: \mathbb{R} \rightarrow \{-1, +1\}$ que él ha diseñado. Para ello le proporciona 100 valores de dicha función y una clase finita de funciones H y le pide que diga que función $h \in H$ aproxima mejor a f . Explique de forma concisa su contestación.

6. Considere de nuevo el escenario descrito para averiguar la función f del matemático. Si el matemático le permitiera añadir una hipótesis extra al problema. ¿Añadiría algo? ¿Qué y por qué?

7. En regresión hemos introducido la función de error cuadrática $e(x)=(h(x)-y)^2$ entre predicción y etiqueta. Sin embargo medidas aberrantes en los valores de las etiquetas producen valores de error enormes que condicionan la búsqueda de buenas soluciones. Con objeto de rebajar la influencia de estos errores aberrantes podemos decidir usar el valor absoluto de la diferencia en lugar del cuadrado como medida de error, es decir $e(x)=|h(x)-y|$. Suponga que dispone de una muestra de N valores unidimensionales $z=\{z_1 \leq z_2 \leq \dots \leq z_N\}$. Calcule que función de la muestra $h(z)$ alcanza el mínimo del error $\text{Error}=\sum_i |h(z)-z_i|$ y argumente a la vista del estimador que le salga si el cambio de norma para eliminar observaciones aberrantes es una buena o mala decisión.

8. Considere un modelo de "bin" para una hipótesis h que comete un error μ al aproximar una función determinística f . Ambas funciones se suponen binarias. Si usamos la misma h para aproximar una versión ruidosa de f dada por

$$P(y|x)=\lambda 1 - \lambda y = f(x)y \neq f(x)$$

¿Qué error comete h al aproximar y , y con qué valor de λ el error de h es independiente de μ ?

9. Consideremos un problema de clasificación binaria probabilístico con etiquetas $\{+1, -1\}$ donde la solución probabilística sin costes esta dada por la función $g(x)=P[y=+1|x]$. Suponga que su problema tiene asociada una matriz de coste que está dada por

Costes

	Verdadero +1	Verdadero -1
decisión: +1	0	10
decisión: -1	1000	0

Identifique la regla de clasificación asociada a $g(x)$. A continuación incorpore los costes a dicha función y deduzca una nueva regla de clasificación que incorpore la información de los costes. Expresar con claridad los pasos que vaya dando y las hipótesis que haga para calcular la solución. ¿Cómo influyen los pesos en la regla final?

10. Es bien conocido que la investigación en "Machine Learning" ha ido produciendo paulatinamente distintos algoritmos de considerable éxito para aproximar problemas de clasificación: k-nn, RL, Árboles, SVM, RRNN, AdaBoost, etc. Sin embargo en los últimos años la redes CNN han mostrado una increíble superioridad frente a las técnicas anteriores en multitud de problemas. Si le piden, como experto, que elija una técnica como la mejor entre las mencionadas, ¿cúal elegiría y por qué?

k-NN: k vecinos más cercanos, RL: Regresión logística., RRNN: redes neuronales, SVM: Maquinas de vectores de Soporte, CNN: redes neuronales convolucionales

11. Suponga un problema de predicción bien definido que hace uso de una clase de funciones H . Suponga que dicha clase H es suficiente como para hacer que $E_{in} \rightarrow 0$ para cualquier muestra de un tamaño dado N . Analice las fuentes del error en el problema de

sobreajuste de esta situación y su evolución cuando se va reduciendo la complejidad de la clase H .

12. Considere $E_n(w) = \max(0, 1 - y_n w^T x_n)$. Muestra que $E_n(w)$ es una cota superior para $[\text{sign}(w^T x_n) \neq y_n]$. Por tanto $1/N \sum E_n(w)$ es una cota superior para E_{in} .