

Trabajo.3: Preguntas de Teoría

Fecha de entrega: 26 Mayo 2019. Valor máximo: 12.5 puntos

NORMAS DE DESARROLLO Y ENTREGA DE TRABAJOS

Para este trabajo como para los demás es obligatorio presentar un informe escrito (hacerlo en pdf, MS Word). **Sin este informe se considera que el trabajo NO ha sido presentado.**

Normas para el desarrollo de los Trabajos: EL INCUMPLIMIENTO DE LAS NORMAS (*) SIGNIFICA PERDIDA DE 2 PUNTOS POR CADA INCUMPLIMIENTO.

- En su informe de contestación debe incluir todas las preguntas en el orden y tal y como se les formula en este documento. (*)
- Las contestaciones irán a continuación de cada pregunta, dejando en blanco las que no conteste. (*)
- Todas las contestaciones deben ser justificadas con argumentos. Sin argumentos la pregunta se considera no contestada.
- Todas las justificaciones matemáticas deben contener todos y cada uno de los pasos de la misma. En caso de duda la contestación no se considerará válida.
- **Forma de entrega:** Subir el pdf a la web de Turnitin.

Todas las preguntas tienen el mismo valor

1. ¿Podría considerarse Bagging como una técnica para estimar el error de predicción de un modelo de aprendizaje?. Diga si o no con argumentos. En caso afirmativo compárela con validación cruzada.
2. Considere que dispone de un conjunto de datos linealmente separable. Recuerde que una vez establecido un orden sobre los datos, el algoritmo perceptron encuentra un hiperplano separador iterando sobre los datos y adaptando los pesos de acuerdo al algoritmo

Algorithm 1 Perceptron

```
1: Entradas:  $(\mathbf{x}_i, y_i), i = 1, \dots, n$ ,  $\mathbf{w} = 0$ ,  $k = 0$ 
2: repeat
3:    $k \leftarrow (k + 1) \bmod n$ 
4:   if  $\text{sign}(y_i) \neq \text{sign}(\mathbf{w}^T \mathbf{x}_i)$  then
5:      $\mathbf{w} \leftarrow \mathbf{w} + y_i \mathbf{x}_i$ 
6:   end if
7: until todos los puntos bien clasificados
```

Modificar este pseudo-código para adaptarlo a un algoritmo simple de SVM, considerando que en cada iteración adaptamos los pesos de acuerdo al caso peor clasificado de toda la muestra. Justificar adecuadamente/matematicamente el resultado, mostrando que al final del entrenamiento solo estaremos adaptando los vectores soporte.

3. Considerar un modelo SVM y los siguientes datos de entrenamiento: Clase-1: $\{(1,1), (2,2), (2,0)\}$, Clase-2: $\{(0,0), (1,0), (0,1)\}$
 - a) Dibujar los puntos y construir por inspección el vector de pesos para el hiperplano óptimo y el margen óptimo.
 - b) ¿Cuáles son los vectores soporte?
 - c) Construir la solución en el espacio dual. Comparar la solución con la del apartado (a)
4. ¿Cuál es el criterio de optimalidad en la construcción de un árbol? Analice un clasificador en árbol en términos de sesgo y varianza. ¿Que estrategia de mejora propondría?
5. ¿Cómo influye la dimensión del vector de entrada en los modelos: SVM, RF, Boosting and NN?.
6. El método de Boosting representa una forma alternativa en la búsqueda del mejor clasificador respecto del enfoque tradicional implementado por los algoritmos PLA, SVM, NN, etc. a) Identifique de forma clara y concisa las novedades del enfoque; b) Diga las razones profundas por las que la técnica funciona produciendo buenos ajustes (no ponga el algoritmo); c) Identifique sus principales debilidades; d) ¿Cuál es su capacidad de generalización comparado con SVM?
7. Discuta pros y contras de los clasificadores SVM y Random Forest (RF). Considere que SVM por su construcción a través de un problema de optimización debería ser un mejor clasificador que RF. Justificar las respuestas.

8. ¿Cuál es a su criterio lo que permite a clasificadores como Random Forest basados en un conjunto de clasificadores simples aprender de forma más eficiente? ¿Cuales son las mejoras que introduce frente a los clasificadores simples? ¿Es Random Forest óptimo en algún sentido? Justifique con precisión las contestaciones.
9. En un experimento para determinar la distribución del tamaño de los peces en un lago, se decide echar una red para capturar una muestra representativa. Así se hace y se obtiene una muestra suficientemente grande de la que se pueden obtener conclusiones estadísticas sobre los peces del lago. Se obtiene la distribución de peces por tamaño y se entregan las conclusiones. Discuta si las conclusiones obtenidas servirán para el objetivo que se persigue e identifique si hay algo que lo impida.
10. Identifique que pasos daría y en que orden para conseguir con el menor esfuerzo posible un buen modelo de red neuronal a partir una muestra de datos. Justifique los pasos propuestos, el orden de los mismos y argumente que son adecuados para conseguir un buen óptimo. Considere que tiene suficientes datos tanto para el ajuste como para el test.

BONUS: Los BONUS solo serán tenidos en cuenta si en el cuestionario obligatorio se ha conseguido al menos un 75 % de los puntos totales.

1. (1.5 puntos) Suponga que durante 5 semanas seguidas, recibe un correo postal que predice el resultado del partido de futbol del domingo, donde hay apuestas substanciosas. Cada lunes revisa la predicción y observa que la predicción es correcta en todas las ocasiones. El día de después del quinto partido recibe una carta diciendole que si desea conocer la predicción de la semana que viene debe pagar 50.000€. ¿Pagaría?
 - a) ¿Cuántas son las posibles predicciones gana-pierde para los cinco partidos?
 - b) Si el remitente desea estar seguro de que al menos una persona recibe de él la predicción correcta sobre los 5 partidos, ¿Cual es el mínimo número de cartas que deberá de enviar?
 - c) Después de la primera carta prediciendo el resultado del primer partido, ¿a cuantos de los seleccionados inicialmente deberá de enviarle la segunda carta?
 - d) ¿Cuántas cartas en total se habrán enviado después de las primeras cinco semanas?
 - e) Si el coste de imprimir y enviar las cartas es de 0.5€ por carta, ¿Cuanto ingresa el remitente si el receptor de las 5 predicciones acertadas decide pagar los 50.000€?
 - f) ¿Puede relacionar esta situación con la función de crecimiento y la credibilidad del ajuste de los datos?
2. (1.5 puntos) Considere un modelo de red neuronal con dos capas totalmente conectadas: d unidades de entrada, n_H unidades ocultas y c unidades de salida. Considere la función de error definida por $J(\mathbf{w}) \equiv \frac{1}{2} \sum_{k=1}^c (t_k - c_k)^2 = \frac{1}{2} \|\mathbf{t} - \mathbf{z}\|^2$, donde el vector \mathbf{t} representa los valores de la etiqueta, \mathbf{z} los valores calculados por la red y \mathbf{w} los pesos de la red. Considere que las entradas a la segunda capa se calculan como $z_k = \sum_{j=0}^{N_H} y_j w_{kj} = \mathbf{w}_k^t \mathbf{y}$ donde el vector \mathbf{y} representa la salida de la capa oculta.
 - a) Deducir con todo detalle la regla de adaptación de los pesos entre la capa oculta y la salida.
 - b) Deducir con todo detalle la regla de adaptación de los pesos entre la capa de entrada y la capa oculta.

Usar θ para notar la función de activación.