

## CSI, examen de aprendizaje automático

17 de diciembre del 2024. Tiempo: 1 hora 50 minutos

El examen es sobre 10.5 puntos.

La respuestas con paja serán penalizadas. Dicho en la terminología de la asignatura: aquellas respuestas en las que el estudiante intenta maximizar recuerdo (*recall*) sacrificando precisión (*precision*) serán penalizadas.

### Pregunta 1 (1 punto)

Queremos categorizar las vocalizaciones que produce un mamífero marino cuyo sistema de comunicación no hemos conseguido descifrar todavía. Podemos usar métodos supervisados y no supervisados. ¿Cuáles son las ventajas de los métodos no supervisados para esta aplicación?

#### Respuesta:

- Al no trabajar sobre un conjunto de datos etiquetados se evitan sesgos antropocéntricos.
- Se ahorra el coste de tener que anotar los datos manualmente.

### Pregunta 2 (1 punto)

Métodos de reducción de dimensionalidad.

1. Menciona un método que sea lo último (*state-of-the-art*) en reducción de dimensionalidad (que no sea t-SNE).
2. Indica cómo se pueden clasificar los métodos de reducción de dimensionalidad vistos en el curso (indica los tipos y ejemplos de cada tipo).

#### Respuesta:

1. Uniform Manifold Approximation and Projection (UMAP).
2. Respecto el tipo de transformación del punto original en un punto en el nuevo espacio de baja dimensión, los métodos se pueden clasificar en lineales (PCA) y no lineales (t-SNE, UMAP).

### Pregunta 3 (1 punto)

Regresión logística.

1. ¿Por qué en la regresión logística no se maximiza la verosimilitud (likelihood) sino el logaritmo de la verosimilitud (log-likelihood)?
2. Al definir la verosimilitud (likelihood) como

$$L(a, b) = \prod_i p(y^i | x^i; a, b),$$

¿cuál es la asunción estadística sobre los datos?

#### Respuesta:

1. Para evitar problemas numéricos. La verosimilitud es normalmente un número muy pequeño y por tanto difícil de representar en un ordenador. Su logaritmo es más fácil de representar.

## 2. Independencia estadística (mutua) entre las instancias.

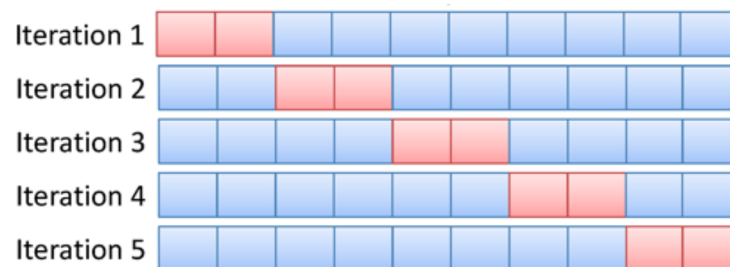
### Pregunta 4 (1.5 punto)

Validación cruzada en  $k$  pliegues ( $k$ -fold cross validation).

1. ¿Qué vale  $k$  en el diagrama de validación cruzada de la figura siguiente?

We split the input data into  $k$   **folds** . Typical value for  $k$  is 10.

At each iteration, the  **blue folds**  are used for  **training** , and  **red folds**  are used as  **validation**



2. Considerad el siguiente fragmento de código.

```
kf = KFold(n_splits=10, shuffle=True)
predictions = cross_val_predict(model, df, y, cv=kf)
```

¿Qué significa el parámetro “shuffle” en la llamada a KFold? ¿Qué sucedería si “shuffle=False”?

3. En el mismo código, ¿qué efecto tendría especificar el parámetro “random\_state=75” en la llamada a KFold?

### Respuesta:

1.  $k = 5$ .
2. “shuffle” indica si las instancias para cada pliegue deben seleccionarse al azar (“shuffle=True”) o bien seleccionando instancias para cada pliegue de forma consecutiva (“shuffle=False”).
3. La selección sería “al azar” pero el resultado siempre sería el mismo, independientemente de la ejecución.

### Pregunta 5 (1 punto)

1. Si lo que realmente importa para determinar la calidad de un modelo es el error de validación, ¿por qué también medimos el error de entrenamiento?
2. ¿En qué condiciones concretas puede suceder que el error de validación y el de entrenamiento de un modelo sean muy parecidos?

### Respuesta:

1. Para saber si hay subajuste o infraajuste.
2. Algunas situaciones  **concretas**  son

- a) Cuando no se han separado correctamente los datos de entrenamiento de los de validación (produciéndose pérdida o *leakage*).
- b) Cuando el modelo es el mismo que ha generado los datos y tenemos suficientes datos para entrenamiento y validación.
- c) Cuando el sesgo del modelo es muy elevado (por ejemplo, un modelo que produce respuestas al azar).

**Pregunta 6** (2.5 puntos)

La declaración de Barcelona para el desarrollo y uso adecuado de la inteligencia artificial en Europa.

1. ¿Cuales son las dos grandes ramas de la inteligencia artificial (IA) y en qué se diferencian?
2. Indica a qué rama pertenece esta parte del curso de CSI.
3. El aprendizaje neuronal profundo, ¿a qué rama pertenece?
4. ¿En qué rama de la IA es más difícil asegurar la fiabilidad y la seguridad?
5. Para asegurar la fiabilidad de la IA, ¿qué considera crucial la declaración de Barcelona?

**Respuesta:**

1. La IA basada en conocimiento (Knowledge-based AI) y la IA basada en datos (data-driven AI). La primera sigue una aproximación descendente (top-down) y hace énfasis en la representación del conocimiento y el razonamiento. La segunda sigue una aproximación ascendente (bottom-up), se basa en el aprendizaje estadístico y requiere grandes volúmenes de datos.
2. A la IA basada en datos.
3. A la IA basada en datos.
4. En la IA basada en datos.
5. Es necesario crear agencias de certificación (verificación y validación) de la IA tal como sucede con medicamentos y otras tecnologías.

**Pregunta 7** (1 punto)

Redes neuronales

1. Indica las partes de una neurona y su función.
2. ¿Qué relación hay entre una neurona artificial y modelos vistos en este curso?

**Respuesta:**

1. La neurona se divide en dendritas (entradas), el cuerpo, que incluye el núcleo de la célula (procesamiento de las entradas) y axones (salidas).
2. El modelo de regresión logística (o un modelo de regresión lineal) puede ser visto como una neurona artificial donde  $y$  sería un axón y las múltiples entradas  $(x_1, \dots, x_i, \dots, x_n)$  serían las dendritas.

**Pregunta 8** (1.5 puntos)

Disponemos de una muestra de valores  $(x, y)$  añadiendo ruido aleatorio a la función coseno en un intervalo en el que la función muestra dos máximos locales. Queremos modelar dichos datos mediante un polinomio de grado  $n$ .

1. Fijamos  $n = 1$  y la regresión produce un modelo que no se ajusta bien a los datos. ¿Por qué?

2. ¿Con qué nombre se conoce a este tipo de error en aprendizaje automático?
3. Observamos que con  $n = 100$  la calidad del modelo mejora. No obstante, si usamos regresión de Lasso, la calidad del modelo empeora al aumentar el parámetro  $\alpha$  de Lasso. ¿Por qué? Usa terminología de aprendizaje automático en tu respuesta.

**Respuesta:**

1. Con  $n = 1$  el polinomio se convierte en una recta. Una recta no es un buen modelo para una función periódica.
2. Sesgo o infraajuste.
3. Al aumentar el valor de  $\alpha$  suficientemente, aumenta el número de coeficientes del polinomio que son cero o tienen un valor muy bajo, lo cual aumenta el sesgo del modelo.