

## Cuestionario NN-SVM, AFI 2022

Nombre:

Apellidos:

### Instrucciones de entrega:

Se esperan respuestas breves que, en promedio, no deberían sobrepasar media cara por pregunta. Contestad bien en el Word o en un pdf, siendo el nombre del archivo a entregar de la forma

apellido\_1\_apellido\_2.xxx

con vuestros apellidos y siendo xxx la extensión. Si las respuestas se escriben a mano, escanead cada hoja, concatenadlas y **entregadlas en un único pdf.**

### Cuestiones:

1. Estoy construyendo un modelo de regresión logística y me preocupa que esté haciendo overfit. ¿Cómo podría verlo en la función de coste?
2. Vamos a aplicar un modelo de 5 vecinos próximos en un problema de clasificación en 2 clases. Indicar los posibles valores que tomarán las probabilidades a posteriori de cada clase.
3. ¿Cómo podemos identificar visualmente las variables más relevantes en un problema de regresión? ¿Y en uno de clasificación?
4. ¿Qué es la maldición de la dimensionalidad? ¿Qué modelos de los estudiados en el curso se van a ver más o menos afectados por ella?
5. Tengo un problema de regresión con muchas features por lo que creo que un modelo SVR **lineal** me puede dar buenos resultados. ¿Qué hiperparámetros debo optimizar? Me da la impresión de que el modelo va a tener la pega de ser homogéneo. ¿Qué puedo hacer para mitigar esto y tener finalmente un buen modelo?
6. Nos hemos despistado en un problema de clasificación y hemos acabado construyendo un modelo de regresión logística **constante** (esto es, en vez de  $w \cdot x + w_0$ , el modelo solo usa  $w_0$ ). ¿Cuánto valdrá  $w_0$ ? ¿Qué predecirá el modelo? (Sugerencia: hacer  $w = 0$  en la fórmula de la log likelihood en las transparencias de la regresión logística y obtener el  $w_0$  que la maximiza.)
7. Estamos trabajando con un modelo SVR **gaussiano** para una muestra de dimensión 2 pero no nos hemos dado cuenta de que las dos features son exactamente iguales. ¿Cómo afectará esto al modelo? ¿Podemos controlar su efecto? ¿Cómo?
8. Hemos entrenado una SVM en un problema de clasificación con una constante  $C=10$  sobre una muestra de 1000 patrones y hemos obtenido 200 vectores soporte, de los cuáles 100 tienen coeficientes alfa con el valor  $C=10$ . ¿Cuál sería la accuracy mínima de dicho modelo?
9. Si a continuación aplicamos el modelo a los 800 patrones que **no son vectores soporte**, ¿cuál sería la accuracy del modelo sobre los mismos?
10. Si en un problema dado hacer el fit de un MLP de regresión tarda 10 segundos, ¿qué tiempo de ejecución debo esperar si mantengo la arquitectura MLP pero triplico el tamaño de la muestra? ¿Qué tiempo debería esperar como mínimo si se tratara de un modelo SVR?