Gi

|  |  |
| --- | --- |
| Colegio Universitario **IES** *Siglo 21* | |
| EXAMEN PARCIAL 1 | |
| **Materia:** APRENDIZAJE AUTOMÁTICO 2 | **Prof:** Coraglio Federico |
| **Modalidad:** PRESENCIAL |  |

**Reservado para el alumno**

|  |  |
| --- | --- |
| **Alumno**: Gino Sibello | **Carrera: IA** |
| **DNI**: 41820055 | **Fecha:04/10/2022** |

###### Reservado para el docente

|  |
| --- |
| NOTA |
|  |

* **Escala de Valoración**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Puntaje | 1-24 | 25-39 | 40-54 | 55-61 | 62-66 | 67-72 | 73-79 | 80-87 | 88-95 | 96-100 |
| Nota | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Valoración | Logro  No satisfactorio | | | Logro básico | Logro satisfactorio | | Logro  Destacado | Logro  Sobresaliente | | Logro  Excelente |

|  |
| --- |
| DESARROLLO DEL PARCIAL |

**TEÓRICO:**

1. Explique ¿por qué se dice que al agregar un término de Regularización como el de Ridge o el de Lasso es posible disminuir el overfitting? (10%)

Cuando trabajas con funciones polinómicas para pronosticar valores y estas buscando obtener los valores dependientes de cada incógnita con el fin de tener tu modelo, te podes encontrar con un situación en la que las variables dependientes se adaptan demasiado a los datos dados para el entrenamiento esto hace que tu modelo entre en overfitting debido a que este modelo cuando busques predecir datos que no observo va a tener un alto grado de error debido a que no se adapto a la realidad si no a los datos de entrenamiento.

Entonces ahí tenes un problema, un problema que Lasso o Ridge resuelven ya que lo que hacen es restringir el valor que toman estas variables dependientes lo cual lleva a que la función no se adapte tanto a los datos y pueda generalizar de una mejor forma.

1. Explique el fundamento de Support Vector Machines. (10%)

Cuando estas en una situación en la que estas buscando clasificar entre dos posibles valores. Lo que buscas es encontrar cual es la frontera entre ambos, ¿a qué me refiero con frontera? Me refiero la media entre las distancias de los dos puntos que mas cerca se encuentran entre si y que son de diferentes targets. Con esta media calculas una perpendicular que pasa en la media de la distancia, y ahí tenes tu frontera, la misma determina si un dato es de una clasificación o de otra ahora armas dos perpendiculares que se trazan en los extremos de la distancia con esto determinas tu carretera un espacio en el cual si un dato es de un target o de otro es un poco difuso y la forma de determinarlo se hacer con el valor del parámetro C y ahí entra mi respuesta de la 3

1. ¿Qué significa que el coeficiente de regularización “C” de SVM, tome un valor grande. (10%)

Determina el nivel de restrictividad que va haber en el espacio entre los dos puntos mas cercanos

1. Ud tiene un dataset con una **gran** cantidad de variables X y **pocas** observaciones, entonces ¿es posible que el modelo tenga tendencia al underfitting o al overfitting? Explique. (10%)

Va tener tendencia a underfitting

1. En el caso de Gradient Descent ¿qué indican la dirección y sentido del vector gradiente?¿Qué uso le damos a este vector? (10%)

Indican hacia donde crece la derivada del punto que estas analizando y el uso que le damos es crear otro vector totalmente opuesto para buscar el mínimo.

**PRÁCTICO**

1. Crear un archivo en formato jupyter notebook (.ipynb) para desarrollar y **describir** paso a paso el código necesario para entrenar un modelo de **SVM**, utilizando solo **GridSearchCV** para la búsqueda de hiperparámetros, que clasifique lo mejor posible el target del siguiente conjunto de datos:

**DataSet:** hongos-IE1.csv

Este conjunto de datos describe los hongos en términos de sus características físicas. Se clasifican en: venenosas o comestibles (p: poisonous, e: edible).

Este conjunto de datos incluye descripciones de muestras hipotéticas correspondientes a 23 especies de hongos con branquias en la familia Agaricus y Lepiota. Cada especie se identifica como definitivamente comestible, definitivamente venenosa o de comestibilidad desconocida y no recomendada. Esta última clase se combinó con la venenosa. La Guía establece claramente que no existe una regla simple para determinar la comestibilidad de un hongo.

Las variables fueron transformadas con OrdinalEncoder para que sean todas numéricas.

(30%)

1. Mostrar las métricas correspondientes al Accuracy y el F1, además especificar el grado de certeza para 3 desvíos estándar, del mejor modelo entrenado. (10%)
2. Guardar el mejor modelo entrenado con el nombre modelo.sav (5%)
3. Guardar el documento de texto con las respuestas del teórico, el archivo de jupyter y el modelo en un archivo comprimido en formato .zip con el siguiente nombre AA2-IE1-APELLIDO-NOMBRE.zip (reemplazar APELLIDO y NOMBRE) y enviarlo por correo electrónico a [fcoraglio@ies21.edu.ar](mailto:fcoraglio@ies21.edu.ar) (5%)