**UNIVERSIDAD CATÓLICA BOLIVIANA “SAN PABLO”  
CARRERA DE INGENIERÍA DE SISTEMAS  
MAESTRÍA EN CIENCIA DE DATOS – 3da VERSIÓN  
MATERIA: MINERIA DE DATOS I  
Prof. MSc. Oswaldo Figueroa Domejean**

**NOMBRE: RAMON WILDER SERDAN CARDENAS**

**EXAMEN Nro. 3**

1. Explique cual es la importancia del concepto de “Bias & Variance Trade-off” en Ciencia de Datos. Explique con un ejemplo tal relevancia.

El error por efecto del sesgo (bias)s de un modelo es la diferencia entre el valor esperado (valor de la predicción) y el valor real. Un bias muy alto da a entender que se trata de un modelo muy simple y no ha habido un correcto entrenamiento. La varianza de un modelo es el grado de variabilidad de la predicción en función a los datos que usemos en el training.

El problema surge a la hora de encontrar un modelo y debemos decidir entre mejorar uno de estos errores, vale decir: disminuir la varianza implica aumentar el bias o disminuir el bias hacer que la varianza aumente. Un ejemplo sería cuando nuestro modelo predice muy bien con datos de entrenamiento, pero con datos recientes no tiene buen ajuste. Tal es el caso de modelos para predecir la votación de un estrato poblacional específico. Si nos enfocamos sólo en población de cierta edad, perderemos información valiosa cuando se presente información de gente de otro grupo con información no entrenada.

1. Explique por medio de un ejemplo el concepto de Regularization en Regresión Lineal.

Se trata de un ajuste que “regulariza” los estimadores de los coeficientes hacia cero. La técoca Ridge y Lasso buscan que los coeficientes tiendan a cero y que tengan un buen ajuste medidos por la suma de cuadrados residual para que este sea pequeño. Utilizando esta técnica también nos encontramos con el problema de la pregunta 1 (trade off entre el sesgo y la varianza).

El ejemplo sería cuando nos enfrentamos a un dataset con un accuracy importante y existe la posibilidad de aplicar las técnicas Lasso o Ridge que pueden arrojar una mejor precisión. Sin embargo, los valores podrían tener una varianza alta lo que podría dificultar las interpretaciones.

1. Interprete el concepto de la métrica “lift” en su relación con “support” por medio de un ejemplo en la técnica de las Reglas de Asociación.

En el supuesto de que los eventos 1 y 2 son estadísticamente independientes, Lift se utiliza para cuantificar la frecuencia el antecedente 1 y el consecuente 2 ocurren de manera simultánea. Es decir, si Lift es igual a 1, indica que ese conjunto aparece una cantidad de veces acorde a lo esperado. Si 1 conduce a 2, el valor de Lift es mayor a uno.

Expresado de otra manera, y relacionándola con la medida Support, tendríamos:

Lift (1->2)= Confidence (1->2)/Support (2)

Es decir, Lift como la relación de la proporción de transacciones que contienen el consecuente 2 que también contienen el antecedente 1 y la proporción de transacciones que contienen 1 y 2.