**Quizzes Eco Computacional**

Semana 3 – High Dimensional Inference

Objetivos de aprendizaje:

* Inferencia clásica: Parámetro vs estimador; Sesgo, ECM, Normalidad, Pruebas de hipótesis, pruebas múltiples,
* Big data inference: FDP, FDR

Preguntas:

1. ¿Por qué los estimadores son variables aleatorias?
   1. Porque varía dependiendo de cuantas variables use para estimar el parámetro
   2. Porque varía dependiendo de cuantas observaciones use en mi base de datos
   3. No son variables aleatorias, son un número fijo (i.e beta = -0.5)
   4. **Porque varían dependiendo de la muestra de la que viene mi base de datos**
2. Conforme sube el tamaño de mi muestra, mi estimador:
   1. **Tiene mayor probabilidad de ser igual al parámetro, menor varianza y se distribuye normal alrededor del parámetro**
   2. Tiene mayor probabilidad de ser igual al parámetro, mayor varianza y se distribuye normal alrededor del parámetro
   3. Tiene mayor probabilidad de ser igual al parámetro, menor varianza y no se distribuye normal alrededor del parámetro
   4. Tiene menor probabilidad de ser igual al parámetro, menor varianza y se distribuye normal alrededor del parámetro
3. Intuitivamente, el estadístico F calcula:
   1. **La proporción de varianza total de la variable y que es explicada por la regresión vs la varianza no explicada por la regresión, ajustando por cuantas variables incluí en la regresión**
   2. La proporción de varianza total de y que es explicada por la regresión vs la varianza total de la variable y, ajustando por cuantas variables incluí en la regresión
   3. La proporción de varianza total de y que no es explicada por la regresión vs la varianza explicada por la regresión, ajustando por cuantas variables incluí en la regresión
4. En Big data, ¿por qué las pruebas de hipótesis simples y múltiples arrojan muchos falsos positivos?
   1. **Porque los estadísticos T y F son crecientes al número de observaciones en una regresión; eso hace que salgan muchas pruebas significativas**
   2. Porque los estadísticos T y F son crecientes al número de variables incluidas en una regresión; eso hace que salgan muchas pruebas significativas
   3. Porque los estadísticos T y F son decrecientes al número de variables incluidas en una regresión; eso hace que salgan muchas pruebas significativas
   4. Porque los estadísticos T y F son decrecientes al número de observaciones en una regresión; eso hace que salgan muchas pruebas significativas
5. Una variable incluida en una regresión de 130 variables tiene un p-value de 0.03. Este es el vigésima variable con menor p-value de las 130. Es un falso positivo para un FDR(0.05)?
   1. **Si**
   2. No

k=20, q=0.05, p=0.03

kq/N = 20\*0.05/130 = 0.007.!(p<=0.007)

Semana 4 – Regresión

Objetivos de aprendizaje:

* Aprendizaje supervisado y no supervisado, in and out of sample, paramétrico y no paramétrico
* Gauss-Markov, omitted variable bias, selección de variables
* Regresión para predicción

Preguntas:

1. Imagina que te piden en tu empresa segmentar a los usuarios en 3 categorías: Mal cliente, cliente regular y buen cliente. ¿Qué tipo de algoritmo parece más útil?
   1. **Supervisado, asumiendo que existe una variable que resuma la calidad del cliente**
   2. **No supervisado, asumiendo que no hay una única variable para medir calidad de cliente.**
   3. Supervisado, asumiendo que no hay una única variable para medir calidad de cliente.
   4. No supervisado, asumiendo que existe una variable que resuma la calidad del cliente
2. Imagina que el modelo poblacional de y es y = beta\_0+beta\_1\*x1 + beta\_2\*x\_2+error. ¿Qué pasa con beta\_1 si estimo la regresión incluyendo x3?
   1. **El sesgo de beta\_1 se mantiene constante, la varianza de beta\_1 se incrementa mientras más correlacionadas estén x3 y x1.**
   2. El sesgo de beta\_1 incrementa, la varianza de beta\_1 se incrementa mientras más correlacionadas estén x3 y x1.
   3. El sesgo de beta\_1 se mantiene constante, la varianza de beta\_1 se disminuye mientras más correlacionadas estén x3 y x1.
   4. El sesgo de beta\_1 se disminuye, la varianza de beta\_1 se incrementa mientras más correlacionadas estén x3 y x1.
3. Con los siguientes datos, estima la varianza de regresión (betas, sigma, se de beta\_1):

(x,y): ({21,24,26,27,29,25,25,30}, {2.8,3.4,3.0,3.5,3.6,3.0,2.7,3.7})

Sigma\_hat = 0.50281931

Se(beta)= 0.06667319

1. En un escenario donde tienes muchísimas observaciones, qué tipo de algoritmos ganan ventaja:
   1. Paramétricos (Menos flexibles)
   2. No Paramétricos (Más flexibles)

Semana 5 – Regularización

Objetivos de aprendizaje:

* Cross Validation: K-fold y LOOCv
* Information Criterias (AIC, AICc, BIC)
* FDR como seleccionador de variables
* Ridge
* Lasso

Preguntas

1. En un K-fold CV, usas el fold K para estimar y los K-1 folds restantes para validar el modelo (T,F)
2. ¿Por qué el FDR podría ser un método de selección de variables inferior a un LASSO?
   1. Porqué si tenemos más columnas que filas, nos quedamos sin grados de libertad y no se puede estimar
   2. Porqué LASSO optimiza directamente por minimizar el ECM y FDR indirectamente
   3. No es un método inferior
   4. Porque LASSO te arroja estimadores sesgados
3. En el LASSO path, escoger el modelo que minimiza el AICc se aproxima \_\_\_, el BIC a \_\_\_ y el AIC a \_\_\_\_
   1. **CV min (el modelo que minimiza el ECM), CV 1se (un modelo con algo de underfitting), y un modelo más complejo y cercano el underfitting**
   2. CV 1se (un modelo con algo de underfitting), un modelo más complejo y cercano el underfitting y CV min (el modelo que minimiza el ECM)
   3. CV min (el modelo que minimiza el ECM), un modelo más complejo y cercano el underfitting, y CV 1se (un modelo con algo de underfitting)
4. ¿En qué escenarios estimarías un ridge o un elastic net sobre un LASSO?
   1. En un escenario donde me interese hacer inferencia y el poder predictivo
   2. En un escenario donde me interese sólo hacer inferencia
   3. **En un escenario donde me interese sólo el poder predictivo**

Semana 6 – Clasificación

Objetivos de aprendizaje:

* Tipos de clasificacion
* Metricas matriz de confusion
* ROC
* Puntos de corte
* PR curve
* Rebalanceo clases: Oversampling, undersampling
* Multiclase, multi label

Preguntas

1. Imaginen que quiero clasificar el tema de un texto de reviews de un producto de Amazon. Los usuarios pueden escribir lo que quieren. ¿Qué tipo de modelo de clasificación es?
   1. Es un modelo de clasificación binaria
   2. Es un modelo de clasificación multiclase
   3. **Es un modelo de clasificación multilabel**
2. Imaginen un modelo donde quiero clasificar fraude de tarjeta de crédito. Si mi modelo detecta fraude, el banco va a bloquear la cuenta del cliente. Ella tendría que ir a la sucursal para poder reactivarla. Si el modelo no detecta el fraude, el dinero será robado y el usuario tendrá que pagar por lo gastado ¿Qué tipo de error y métrica es la principal de mi modelo?
   1. Ambos son graves, pero es peor el falso negativo. Nos interesa más el recall/sensitivity
   2. **Ambos son graves, pero es peor el falso positivo. Nos importa más el precission**
   3. Ambos son graves, pero es peor el falso positivio. Nos importa más el specificity
3. Cuál es la principal dimensión para elegir entre oversampling y undersampling
   1. El tipo de clasificación (binaria, multiclase, multilabel)
   2. El error fuera de la muestra
   3. **El tamaño de la muestra**
4. Observas modelo que tiene un accuracy de 70% pero un AUC ROC de 0.55. ¿Es un modelo listo para ir a producción?
   1. **No, sin importar que tipo de error es más grave; parece es un modelo que no mejora una volado.**
   2. Depende de qué tipo de error nos interese más
   3. Si. El accuracy es suficientemente alto

Semana 7 – Árboles de decisión

Objetivos de aprendizaje:

* Arboles de clasificación
* Arboles de regresión
* Tree pruning
* Boostrapping

Preguntas

1. ¿Si una variable ya fue usada en un nodo interno, puede ser usada de vuelta?
2. ¿Qué significa que un árbol sea greedy?
   1. **Que siempre busca el Split que más disminuye el RSS; sin importar la secuencia de futuros splits.**
   2. Que siempre busca la secuencia de splits que más disminuye el RSS; tomando en cuenta la secuencia de futuros splits.
   3. Que siempre busca el Split que menos disminuye el RSS; sin importar la secuencia de futuros splits.
   4. Que siempre busca el Split que más disminuye el RSS; tomando en cuenta la secuencia de futuros splits.
3. ¿Qué parámetros irías calibrando para estimar un árbol?
   1. **Min obs: la cantidad mínima de observaciones en nodos terminales; Min cut: La cantidad mínima en nodos intermedios, y min dev: La reducción mínima del deviance del modelo**
   2. Min obs: La cantidad mínima en nodos intermedios; Min cut: la cantidad mínima de observaciones en nodos terminales, y min dev: La reducción mínima del deviance del modelo
   3. Min obs: la cantidad mínima de observaciones en nodos terminales; y min dev: La reducción mínima del deviance del modelo
   4. Min obs: la cantidad mínima de observaciones en nodos terminales; Min cut: La cantidad mínima en nodos intermedios, y min dev: La cantidad máxima de error permitido
4. El set de modelos a evaluar vía CV en los árboles es \_\_\_\_\_ comparado con un LASSO porque en el primero evalúo variables y \_\_\_\_\_ en los que hacer cortes

Semana 7 – Bagging, Random Forests, AdaBoost

Objetivos de aprendizaje:

* Bagging
* MTRY
* Boosting
* GBM

Preguntas

1. Un bagging tree es un subcaso de un Random Forest (T o F)
2. Al aplicar muestreo sobre las observaciones que accede cada árbol en Bagging Trees, RF y XGBM, ¿qué parte del error cuadrático intento disminuir?
   1. El sesgo
   2. La varianza
3. ¿Cuáles son las diferencias entre Boosting Machines y Random Forests?
   1. El RF se realiza de manera paralela, el boosting de manera secuencial. En ambos cada árbol intenta aprender algo nuevo de la variable endógena, pero el RF lo hace sin mirar errores previos.
   2. El RF se realiza de manera secuencial, el boosting de manera paralela. En ambos cada árbol intenta aprender algo nuevo de la variable endógena, pero el RF lo hace sin mirar errores previos.
   3. El RF se realiza de manera paralela, el boosting de manera secuencial. En ambos cada árbol intenta aprender algo nuevo de la variable endógena, pero el Boosting lo hace sin mirar errores previos.
   4. El RF se realiza de manera paralela, el boosting de manera secuencial. En ambos cada árbol intenta aprender algo nuevo de la variable endógena, pero el RF lo hace mirando errores previos.
4. El Random Forest es un subcaso de un XGB con colsubsample y sample < 1 (T o F)

Semana 8 –GBMs, XGBoost, K-means, K-mediods and Hierarchical clustering

Objetivos de aprendizaje:

* GBM
* XGBoost
* Kmeans
* Hierarchical

Preguntas

1. En un ADABoost, entre peor sea un árbol, más peso tendrá en el estimador final
2. ¿Para qué sirve el Learning rate hiper-parameter?
   1. **Es un parámetro que obliga a GBM a seguir buscando patrones/construyendo árboles.**
   2. Es un parámetro que pondera el porcentaje de observaciones en cada terminal node, controla overfitting
   3. Es un parámetro que le da peso distinto a las observaciones donde se equivocó el árbol anterior
3. El Random forest es un subcaso del XGBoost aplicando subsample de columnas y filas
4. ¿Porque el K-means es sensible a variables con outliers?
   1. Porqué los centroides se calculan con distancias no estandarizadas
   2. Porqué por las medias son medidas sensibles a outliers
   3. Porque el random start puede ser malo
5. Si conozco el número de clústeres a estimar, típicamente preferiría un Hierarchical clustering vs un K-mediods.