Examen Parcial 3. COM 23701 33701 Aprendizaje de Máquina. Tema: Datos de Muchas Dimensiones

Prof. Marco Morales, Departamento Académico de Computación, ITAM

Otoño de 2022

Este examen parcial tiene 6 preguntas con un máximo de 60 puntos para resolver en 90 minutos. Instrucciones: 0) Imprime este examen para responderlo de tu puño y letra. 1) Escribe abajo tu nombre completo y firma de conformidad con estas instrucciones y con seguir las estipulaciones sobre integridad académica del reglamento de alumnos. 2) Escribe tu clave única en todas las páginas. 3) Responde este examen en forma individual, apoyándote exclusivamente en tus propias notas y en las notas de la clase disponibles en Canvas y sin recibir ningún tipo de ayuda de otras personas ni de asistir a otros alumnos en sus propios exámenes. 4) Responde cada pregunta dentro del recuadro correspondiente. 5) Entrega el examen respondido mediante la interfaz de `gradescope.com`.

Nombre: Alvaro López Zamora Firma: AZ Fecha: 16-11-22 C.U.: 198442

1. (10 puntos) Explica en tus propias palabras que es el *Curse of Dimensionality* y su relación con la distancia entre los datos y con la eficiencia computacional para explorar un espacio.

El Curse of Dimensionality se refiere un fenómeno que ocurre cuando tenemos datos de muchas dimensiones. A lo que se refiere este fenómeno es a que, mientras las dimensiones de los datos aumentan, las distancias entre ellos también aumentan. De hecho, aumentan de manera exponencial. La consecuencia de este fenómeno es que, en cálculos como la media o la covarianza para explorar un espacio, el costo computacional aumenta, es decir, las operaciones se vuelven más caras y se necesitan más datos para obtener estimaciones precisas.

2. (10 puntos) ¿Qué propiedades de los datos deben preservar las transformaciones que se aplican a los datos de muchas dimensiones para facilitar su análisis?, describe de una de estas transformaciones y explica de que manera preserva las propiedades mencionadas.

En general, una transformación de datos siempre debe mantener que la media de los datos sea cero y que la matriz de covarianzas sea el producto de la diagonalización de Σ. Es decir, que la matriz de covarianzas sea Λ. Ahora, un ejemplo de estas transformaciones es el Principal Component Analysis que transforma los datos a una representación de menor dimensionalidad. La manera en que mantiene estas propiedades está en que, como escala los datos sobre el eje de uno de sus componentes principales que viene de Λ, la matriz de covarianzas es Λ. Y, además, como se trasladan, la media es cero.

3. (10 puntos) ¿Qué propiedades tiene la matriz de covarianza de un conjunto de datos de N elementos y d parámetros y cual es su aplicación en *Principal Component Analysis*?

En primer lugar cumple con que es una matriz positiva semidefinida. Esto implica que es una matriz simétrica con todos sus eigenvalores no-negativos, es decir, son mayores o iguales a cero. También cumple que en la diagonal guarda las varianzas de cada dato en el orden que se encuentran en el conjunto.

Ahora bien, en Principal Component Analysis la matriz de covarianza se utiliza para encontrar los eigenvalores λ a través del proceso de diagonalización. Dicho proceso encuentra una matriz U tal que $U^T \Sigma U = \Lambda$, con Λ la matriz de eigenvalores en órden decendiente. Entre mayor sea λ, mayor peso tiene en los datos.

4. (10 puntos) Explica en tus propias palabras cómo, en *Principal Component Analysis*, se calcula el error del modelo que resulta de la selección de s parámetros con respecto al modelo original que tiene d parámetros.

Primero, es importante mencionar que esos s parámetros que se seleccionan corresponden a los s componentes más grandes de Λ y el resto de los componentes se reemplazan con cero. Una vez que se seleccionan, se procede a calcular el error del modelo. El cálculo parte del error medio cuadrático entre los datos y el nuevo data set P. Se hace la sustitución correspondiente para llegar a la suma sobre los componentes de la suma de los cuadrados de los d-s features que, por la división entre N de la expresión, resulta en la suma de las varianzas de los d-s features. Lo anterior corresponde a la suma de los d-s eigenvalores. Por lo tanto, la expresión del error relativo a minimizar resulta ser la razón entre la suma de los d-s eigenvalores sobre la suma de los d eigenvalores:

$$\frac{\sum_{j=s+1}^{d} \lambda_j}{\sum_{j=1}^{d} \lambda_j}$$

5. (10 puntos) Explica en tus propias palabras las diferencias entre *Principal Component Analysis* y *Principal Coordinates Analysis*

Como sabemos, ambos métodos tienen el objetivo de transformar conjuntos de datos de muchas dimensiones a una representación de menor dimensionalidad. Sin embargo, Principal Coordinates Analysis busca preservar la distancia de los datos al hacer la transformación. Además, los conjuntos de datos sobre los que hacen la aproximación de pocas dimensiones son distintos. En el caso de Principal Component Analysis, la aproximación se hace sobre el conjunto X. En el caso de Principal Coordinates Analysis, la aproximación se hace sobre el conjunto XX^T.

6. (10 puntos) ¿Explica para qué sirve el algoritmo NIPALS y en que casos es ventajoso utilizarlo?

El algoritmo NIPALS sirve para suavizar el ruido Gaussiano y suavizar el ruido que pueda venir de cuentas o datos faltantes. Esto lo hace formulando actualizaciones como sumas en lugar de operaciones de matrices e ignora los datos faltantes. Por lo tanto, en general, es ventajoso utilizarlo cuando tenemos un conjunto de datos que pueda tener entradas faltantes.