

TÉCNICA DEL APRENDIZAJE NO SUPERVISADO: REDUCCIÓN DE DIMENSIONES

Nelson Yoel Phuño Cahuana; Daysi Saimira Machaca Condori

INTELIGENCIA COMPUTACIONAL

nelsonyoelpc@gmail.com; zaimira.mlc@gmail.com

REDUCCIÓN DE DIMENSIONES

En los problemas de Machine learning se involucra miles o millones de funciones para cada instancia de entrenamiento, Esto hace que el entrenamiento sea extremadamente lento y también mucho más difícil encontrar una buena solución. En un problema real constantemente es posible reducir la cantidad de funciones de esta manera convertir un problema insoluble en uno manejable. (Géron, 2019)

- El escalamiento multidimensional: reduce la dimensionalidad mientras intenta preservar las distancias entre las instancias.
- Isomap: crea un gráfico conectando cada instancia a sus vecinos más cercanos, luego reduce la dimensionalidad mientras se intenta preservar las distancias entre las distancias
- La incrustación de vecinos estocásticos distribuidos en t (t-SNE): esto reduce la direccionalidad mientras intenta mantener instancias similares cercanas y distintas separadas frecuentemente es utilizado para la visualización de instancias en espacios de alta dimensión,
- El análisis discriminante lineal (LDA) es un algoritmo de clasificación, pero durante el entrenamiento aprende de los ejes más discriminantes entre las clases. Los ejes pueden ser definidos para un hiperplano sobre el que proyectar los datos no obstante los beneficios son que la proyección mantendrá las clases lo más alejadas posible esto permite reducir la dimencionalidad antes de ejecutar otra clasificación.

¿Cómo funciona?

La reducción de dimensiones es frecuentemente usada como una etapa de pre proceso en el entrenamiento de sistemas, y consiste en escoger un subconjunto de variables, de tal manera, que el espacio de características quede óptimamente reducido de acuerdo a un

criterio de evaluación, cuyo fin es distinguir el subconjunto que representa mejor en el espacio inicial de entrenamiento. Como cada característica que se incluye en el análisis, puede incrementar el costo y el tiempo de proceso de los sistemas, hay una fuerte motivación para diseñar e implementar sistemas con pequeños conjuntos de características. Sin dejar de lado, que al mismo tiempo, hay una opuesta necesidad de incluir un conjunto suficiente de características para lograr un alto rendimiento en qué tipo de problemas son más apropiados. (ANDRÉS J. HERNÁNDEZ, 2006)

¿Cómo se mide su eficiencia?

La eficacia se observa al visualizar, imprimir en una gráfica donde el sobre ajuste use las características realmente importantes, cuando tiene una menor intensidad computacional, muchas dimensiones ocasionan que el poder computacional se sobrecargue. Finalmente, cuando se obtiene características independientes adecuados.

¿A qué tipos problemas mejor responde?

- En la categorización de textos, donde cada atributo corresponde a la ocurrencia de una palabra clave o un término clave del documento.
- Recuperación de imágenes, donde se ha producido una explosión en la colección de imágenes y es necesario indexar las imágenes para recuperarlas de forma eficiente.
- La detección de intrusiones en sistemas informáticos

Bibliografía

ANDRÉS J. HERNÁNDEZ, E. D. (2006). REDUCCIÓN DE DIMENSIONES PARA CLASIFICACIÓN DE DATOS MULTIDIMENSIONALES USANDO MEDIDAS DE INFORMACIÓN . (32).

Géron, A. (2019). Hands-on Machine Learning with Scikit-Learn, Keras & TensorFlow. Nicole Tache.

Rodrigo, J. A. (01 de 02 de 2020). Recuperado el 15 de 02 de 2021, de [https://www.cienciadedatos.net/documentos/52_deteccion_anomalias_autoencoder_pca.html#:~:text=La%20detecci%C3%B3n%20de%20anomal%C3%ADas%20\(outliers,no%20anomal%C3%ADa\)%20de%20las%20observaciones](https://www.cienciadedatos.net/documentos/52_deteccion_anomalias_autoencoder_pca.html#:~:text=La%20detecci%C3%B3n%20de%20anomal%C3%ADas%20(outliers,no%20anomal%C3%ADa)%20de%20las%20observaciones).

