

Tarea 1

Septiembre 2021

Profesor: Ángel Jiménez M.

Auxiliar: Rafael De la Sotta V.

Ayudantes: Camila Jorquera, Cristian Correa, Gabriel Iturra, Martín Ríos.

Imagine que se encuentra trabajando en una empresa comercializadora de productos, en la cual los problemas de manejo de inventario están a la orden del día. Ya se ha corrido la voz entre los colegas que usted ha tenido éxito en la resolución de problemas típicos de gestión utilizando herramientas de deep learning. En particular, ha utilizado redes neuronales feedforward para modelar precios de productos, conductas de clientes (fuga y riesgo) y, además, existe el rumor que en su tiempo libre ha logrado clasificar pacientes diabéticos y no diabéticos como aporte a la investigación en salud.

Estos rumores han escalado hasta la gerencia de operaciones, desde donde le piden consejos para la generación de un modelo que responda ciertas dudas generales sobre el inventario de productos, tales como:

- ¿Cuáles son los productos que deberíamos dejar de vender?
- ¿Es posible saber cuántas ventas tendrá cada producto?
- ¿Es posible generar algún tipo de “ranking” de los productos (por ejemplo, en cuanto a ventas), ya sea de los que debemos eliminar o de los que debemos mantener?

Como experto, usted le explica a la gerenta de operaciones que los modelos de deep learning, en general, se presentan como “cajas negras”, y que por lo tanto conocer el impacto de las variables explicativas es difícil aunque posible, pero que usted no tiene tiempo para llegar hasta ese punto en estos momentos, a lo que le responde:

“No nos interesa en esta primera vuelta saber las variables explicativas, solo mejorar el sistema de inventario con algún modelo predictivo en el cual confiar para la temporada que está empezando. Actualmente tenemos modelos predictivos creados con machine learning, pero no son muy buenos. Nos gustaría saber qué opinas luego de usar deep learning, por último, para tener algo contra qué compararnos”.

Su respuesta es la siguiente:

“Parece un problema interesante. No aseguro nada, pero me gustaría echarle una mirada, por favor envíenme la hoja de cálculo con los datos”

En respuesta, obtiene un archivo llamado *inventario.csv* y un detalle de algunas de sus columnas.

- Order: es sólo un indicador de orden, se puede ignorar.
- Filetype: ‘histórico’ (producto del pasado), ‘activo’ (en la actualidad)
- SKU number: identificador único de cada producto.
- SoldFlag: 1 si se vendió en los últimos 6 meses, 0 si no.
- SoldCount: cuantas ventas tuvo el producto en los últimos 6 meses.
- MarketingType: dos tipos de marketing realizados (S = ‘social network’, D = ‘direct marketing’)
- Release Number: número de versión.
- New Release Flag : cualquier producto que haya tenido un lanzamiento nuevo desde la versión 1 (es decir, release number > 1).
- StrengthFactor: valor representativo de resistencia del producto (no se tiene mayor información pero es calculable para cada producto)
- ReleaseYear: año de lanzamiento.
- ItemCount: cantidad de inventario de cada producto.
- PriceReg: precio recomendado de mercado.
- LowUserPrice: precio pagado por los usuarios.
- LowNetPrice: precio pagado a los proveedores.

Resuelva las inquietudes de la gerenta de operaciones por medio de los siguientes pasos:

P1.- (0.5) Para responder la primera pregunta, identifique alguna(s) variable(s) que permita(n) distinguir cuál producto debe ser eliminado y haga un análisis estadístico descriptivo de las variables de los productos que se mantendrán. ¿Qué conclusiones obtuvo?

P2.- (0.5) Pensando en la segunda pregunta - ¿Es posible saber cuántas ventas tendrá cada producto? - describa (de manera escrita, sin implementación) una solución que identifique una variable objetivo y un modelo de pronóstico.

P3.- (1.0) Volviendo a lo planteado en P1, proponga una red neuronal que permita determinar qué productos “activos” se deben retirar del inventario. Explique cómo generar los conjuntos de entrenamiento, validación y test. Note que la variable Filetype dice qué productos están activos y cuáles son registros del pasado.

P4.- (1.5) Implemente la red neuronal propuesta en P3. Haga una selección de hiperparámetros y utilice fundadamente métodos de mejoramiento y regularización. Entregue un diagrama del modelo final.

P5.- (2.0) Aplique la red anterior sobre los datos de inventario activo e identifique cuáles productos deben ser eliminados. Al mismo tiempo responda la pregunta que se le planteó: ¿Es posible generar algún tipo de “ranking” sobre los productos (por ejemplo, en cuanto a ventas), ya sea de los que debemos eliminar o de los que debemos mantener? Si su respuesta es afirmativa, adjunte una hoja de cálculo con los productos activos y su ranking como columna.

P6.- (0.5) Explique por qué es útil enfrentar este problema con mecanismos de deep learning en comparación a métodos tradicionales de machine learning.

P7.- (BONUS de 1.0) Implemente el modelo planteado en P2.

Evaluación de la tarea

La tarea se realiza en grupos. Se debe entregar el código reproducible en un cuaderno de Jupyter o en Google Colab, explicando de manera ordenada los pasos realizados. Se recomienda usar semillas en las funciones random para una evaluación más directa.

Además de la implementación se debe hacer entrega de un informe breve que detalle el procedimiento llevado a cabo en la lista de pasos P1-P6.

Dado que el problema se puede abordar de diferentes maneras, se evaluará la implementación de modelos, utilidad del análisis estadístico, la coherencia del informe y las respuestas entregadas según los resultados obtenidos.

Recuerden que el foro siempre está abierto para preguntas.