

Optimización con SGD vs el algoritmo ADAM

Sebastián Sabat, Rocío Yáñez.

Profesor: Joaquin Fontbona T.

Auxiliares: Álvaro Márquez; Matias Ortiz; Catalina Lizana

Modelo Predictivo para resultados PAES 2022

El objetivo de esta investigación es entregar un predictor del rendimiento de estudiantes en la prueba específica de matemáticas. Para ello, se trabajó con una base de datos que contenía los resultados en todas las pruebas y características sociodemográficas de los estudiantes de la promoción. El trabajo se dividió en dos etapas, la primera consistió en determinar qué atributos explicaban mejor la variable de interés, para ello se estudió la correlación de Pearson y se realizaron test de hipótesis. La segunda etapa fue implementar los algoritmos ADAM y SGD.

De manera preliminar, se buscó filtrar la base de datos con el objetivo de reducir sesgos de selección, para lo cual se consideraron sólo a los estudiantes que cumplieren:

- Haber realizado las pruebas obligatorias y la prueba electiva, además de la prueba específica de matemáticas.
- Ser estudiante de planes de estudio convencionales (enseñanza H-C o formación técnica profesional).
- Haber egresado en el año 2022 del colegio y nacido el 2004.

Estas restricciones permiten descartar casos patológicos de la data.

Implementación

Se trabaja con un modelo de clasificación binaria, donde se considera un corte de 500 puntos que indica si al estudiante le fue bien o mal. Se implementó para ello el entrenamiento de una red neuronal. Optimizando la función Binary Cross Entropy.

La data se normalizó y se balanceo, para esto se realizó un submuestreo aleatorio. Se implementó la regularización L2 para evitar el overfitting.

En ambos casos se elige un learning rate inicial de 0.01, junto un batch size de tamaño 20. Se realizan experimentos variando estos y otros parámetros

SGD vs ADAM

El método del SGD consiste en optimizar una función objetivo a partir de aproximar su gradiente real a partir

de una colección de puntos iid's que se ocupan como input para estimar dicho gradiente.

Luego se prosiguió con el algoritmo Adaptive Momentum (ADAM), consiste principalmente en dos componentes: descenso de gradiente de minibatch y programas de tasa de aprendizaje. Esto permite que el algoritmo se adapte más rápidamente a los cambios en los datos de entrenamiento. Los programas de tasas de aprendizaje son conjuntos de learning rates para diferentes capas de la red. Determinan qué tan rápido la red aprenderá a reconocer patrones en función de los datos de entrenamiento que recibe.

Resultados

Si bien ambos métodos entregan soluciones similares, presentan algunas diferencias, por ejemplo en la cantidad de épocas que necesita cada uno para alcanzar una aproximación similar, donde el algoritmo ADAM es el que requiere un menor tiempo de entrenamiento.

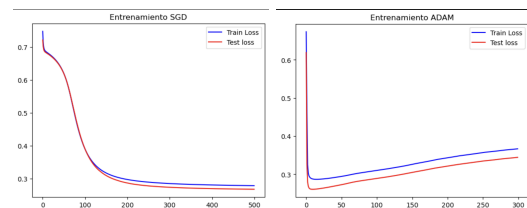


Figura 1: Train Loss y Test Loss usando Descenso de Gradiente Estocástico y ADAM

Referencias

- [1] Mehmood, F.; Ahmad, S.; Whangbo, T.K. An Efficient Optimization Technique for Training Deep Neural Networks. Mathematics 2023, 11, 1360. <https://doi.org/10.3390/math11061360>
- [2] Bottou, L. (1998). On-line learning and stochastic approximations. In On-Line Learning in Neural Networks, 9–42. Ver [enlace](#)
- [3] Centro de estudios, Mineduc, Datos abiertos. PAES 2023, Prueba de Acceso a la Educación 2023 - Inscritos Puntajes. Disponible [aquí](#).