
Bayesian Posterior Sampling via Stochastic Gradient Fisher Scoring

MA4402-1 Simulacion Estocástica: Teoría y Laboratorio

Branco Paineman, Vicente Cabezas^{* 1}

Abstract

Los algoritmos de machine Learning, junto con el avance de la complejidad de las arquitecturas, necesitan la implementacion de nuevos metodos de cálculo de parametros, en la respectiva presentacion ahondaremos en particular en un método bayesiano para el cálculo de los parámetros de una red neuronal simple por medio de el algoritmo Stochastic Gradient Fisher Scoring (SGFS) y compararemos sus resultados en un experimento de reconocimiento de digitos con la variante estandar de SGD, el objetivo de esta investigacion es comparar experimentos realizados en.

1 Modelamiento y problema

El problema se modela mediante el uso de librerias especializadas en Deep Learning, en este caso Pytorch, donde se construye una red neuronal simple de poca , con una función de activación sigmoid, con el objetivo de generar una clasificación binaria sobre imágenes en blanco y negro en un data set de numeros del 1 al 10, el cual fue filtrado con el objetivo de realizar el entrenamiento solo sobre los numeros 9 y 7.

1.1 Idea del algoritmo y datos

El algoritmo SGFS posee la intencion de calcular dado una distribucion a priori, los parametros "óptimos" de un algoritmo de ML. El algoritmo SGFS proviene de la implementacion del algoritmo SGLD, el cual es una variante estocastica del algoritmo de descenso de gradiente SGD, la principal caracteristica de SGFS es que por medio del teorema central del limite (Para el caso bayesiano) extiende SGLD donde para tasas grandes de avance (ϵ grande) aproximara la distribucion posterior mediante una distribucion normal y para tasas pequeñas imitara el comportamiento de SGLD. Cabe destacar que uno de los puntos más importantes de este algoritmo que lo hace más eficiente, es que se utilizan solamente mini-batches en cada iteración en vez de todo el dataset.

Los datos fueron obtenidos de la base de MNIST, que puede ser descargada desde pytorch, la base corresponde

$$\theta_{t+1} \leftarrow \theta_t + \frac{\epsilon C}{2} \{ \nabla \log p(\theta_t) + N \bar{g}_n(\theta_t; X_n^t) \} + \nu$$

where $\nu \sim \mathcal{N}(0, \epsilon C)$

Figure 1: Ecuación utilizada en el Algoritmo SGLD, donde ϵ es la tasa de avance, C la matriz preconditionante y ν es una variable aleatoria que representa el ruido Gaussiano.

a un conjunto de datos de 70.000 imágenes cuadradas pequeñas en escala de grises de 28×28 píxeles de un solo dígito escritas a mano entre 0 y 9.

Los datos fueron filtrados con el objetivo de realizar una clasificación binaria sobre los numeros 7 y 9.

$$\theta_{t+1} \leftarrow \theta_t + \frac{1}{2 \left(\gamma N \hat{I}_{1,t} + \frac{4B}{\epsilon} \right)^{-1}} \{ \nabla \log p(\theta_t) + N \bar{g}_n(\theta_t) + \eta \}$$

Figure 2: Ecuación del Algoritmo SGFS, donde ϵ es la tasa de avance, B es cualquier matriz semi-definida positiva, η es una variable aleatoria que representa el ruido, γ es el tamaño de los mini-batch, $\hat{I}_{1,t}$ es información de Fisher

1.2 Presentación

En la presentación se mostraran los resultados y la comparación del modelo SGFS con el modelo estándar de SGD, esto mediante métricas estándar para la comparación de modelos de ML.

Referencias

- [1] Sungjin Ahn, Anoop Korattikara, Max Welling (2012). Bayesian Posterior Sampling via Stochastic Gradient Fisher Scoring. <https://arxiv.org/ftp/arxiv/papers/1206/1206.6380.pdf>