

# Práctica N° 1

Thomás Coronel

*Redes Neuronales, Instituto Balseiro, Bariloche, Argentina*

*E-mail: thomas.coronel@ib.edu.ar; thomascoroneltuc@gmail.com*

Corrige: Ariel Curiale

En el presente informe se realiza la implementación en el programa Python de sistemas de aprendizajes supervisados y no supervisados. Se trabajó con la base de datos CIFAR-10 y MNIST clasificando las imágenes de cada set de datos, midiendo la exactitud del método Suport Vector Machine (*SVM*), Soft Max (*SfM*) y el costo de los mismos. También se utilizó el método de k-nearest neighbors para clasificar un set de datos generados de manera aleatoria.

## I. EJERCICIO 1

En éste ejercicio se generó u conjunto de datos de 5000 datos de dimensión  $n = 5$ . Se realizó la implementación de regresión lineal utilizando el método de mínimos cuadrados. En la figura 1 se muestra una de las dimensiones y la recta de ajuste de la misma. El error obtenido en los parámetros es del 3.642 %.

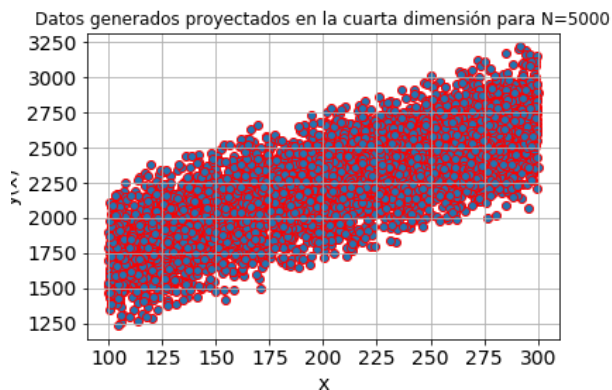


Figura 1: Set de datos de 5000 muestras de dimensión  $n=5$  generados.

## II. EJERCICIO 2

Se generó un conjunto de datos de dimensión  $n = 3$  de 5 clases cada una de 5000 muestras. Se clasificó para k-nearest neighbors de  $k=1$  hasta  $k=7$ .

En el presente informe no se realiza la muestra de cómo fue la evolución de la clasificación del método debido a problemas en el código realizado. En un futuro cercano se proveerá dichos resultados.

## III. EJERCICIO 3

Se implementó el método de clasificación lineal k-nearest neighbors para clasificar las imágenes de la base de datos MNIST y CIFAR-10.

Para el caso de MNIST se tomó una iteración de 1 a 7 vecinos con un número de test de 50 datos llegandose en todos los casos a una precisión del 32 %.

Para el caso de CIFAR-10 se tomó una iteración de 1 a 5 vecinos con un número de

test de 30 datos llegándose en todos los casos a una precisión de 30 %.

#### IV. EJERCICIO 4

Se realizó un set de datos aleatorios, los cuales contaban con 5 clases de 1000 muestras cada uno de dimensión  $n=3$ . Se realizó el método k-nearest neighbors para clasificar los datos tomando  $k=1$  hasta  $k=7$  obteniéndose una exactitud del 100 % en todos los casos realizados.

#### V. EJERCICIO 5 Y 6

En éstos ejercicios se muestra cual es la exactitud de clasificar las bases de datos MNIST y CIFAR-10 mediante dos métricas distintos *SVM* y *SfM*.

##### A. MNIST

En la presente sección se muestra cómo fue el costo y la exactitud para el set de datos de MNIST para el método *SVM* y *SfM*. Se realizó mediante  $epocas = 300, batchsize = 200, lr = 1e - 9$  y  $reg = 1e - 5$ .

En las figuras 2 y 3 se muestra el costo de haber realizado la clasificación mediante *SVM* y *SfM* respectivamente. Se puede observar que el costo de *SVM* es mayor al de *SfM* para las primeras 50 épocas, luego esto cambia, debido a que *SVM* disminuye a un costo menor a 1 mientras que para el final de las 300 épocas *SfM* tiene un costo aproximado de 1,4. Por lo que se concluye que el método *SVM* para éste problema es mejor.

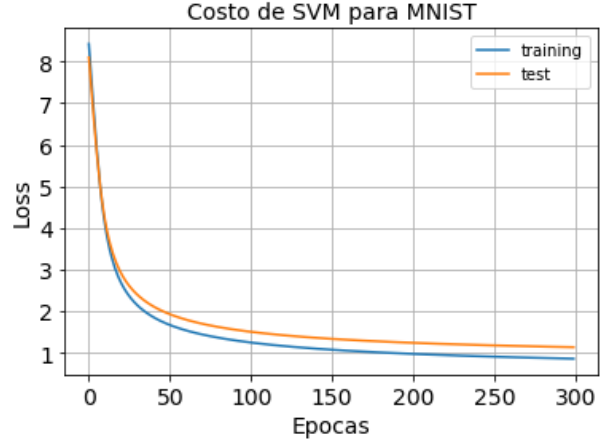


Figura 2: Costo de clasificación mediante el método SVM para el set de datos MNIST.

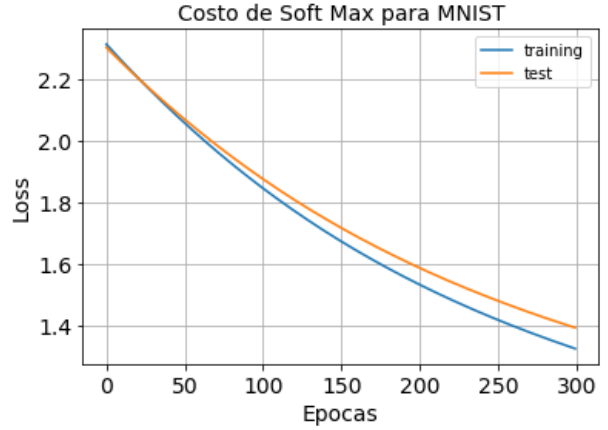


Figura 3: Costo de clasificación mediante el método SfM para el set de datos MNIST.

En las figuras 4 y 5 se muestran muestra la exactitud en la clasificación mediante *SVM* y *SfM* respectivamente. Se observa que para las 300 épocas que se realizaron el *SfM* presenta una menor exactitud, aproximadamente 10 % por lo que éste método es menos eficiente. Caso contrario *SVM* para obtener el mismo resultado le requirió aproximadamente 35 épocas.

Como conclusión *SVM* es menos costoso a partir de una cierta cantidad de épocas y tiene una mayor exactitud en la clasificación. En

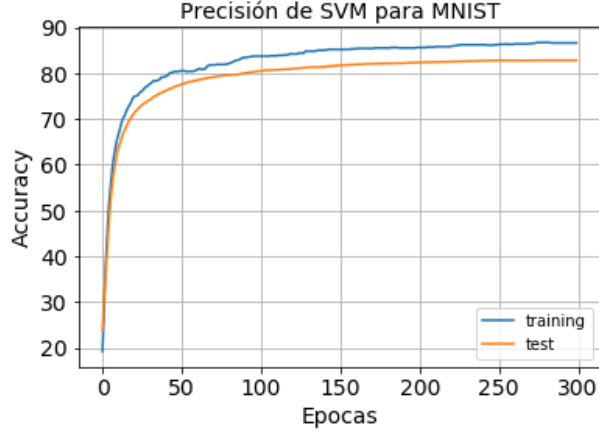


Figura 4: Exactitud de clasificación mediante el método SVM para el set de datos MNIST.

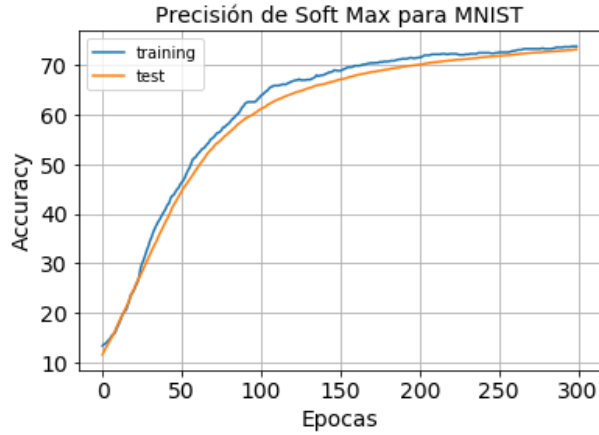


Figura 5: Exactitud de clasificación mediante el método SfM para el set de datos MNIST.

*SfM* es menos costoso en un cierto rango, pero no llega a alcanzar la exactitud de *SVM*.

## B. CIFAR-10

En la presente sección se muestra cómo fue el costo y la exactitud para el set de datos de CIFAR-10 para el método *SVM* y *SfM*. Se realizó mediante  $epocas = 200, batchsize = 200, lr = 1e - 8$  y  $reg = 1e - 5$ .

En las figuras 6 y 7 se muestra el costo de haber realizado la clasificación mediante *SVM* y *SfM* respectivamente. Se puede observar que el costo de *SVM* es mayor al de *SfM* en todas las épocas realizadas. Por lo que el método de *SfM* es más conveniente.

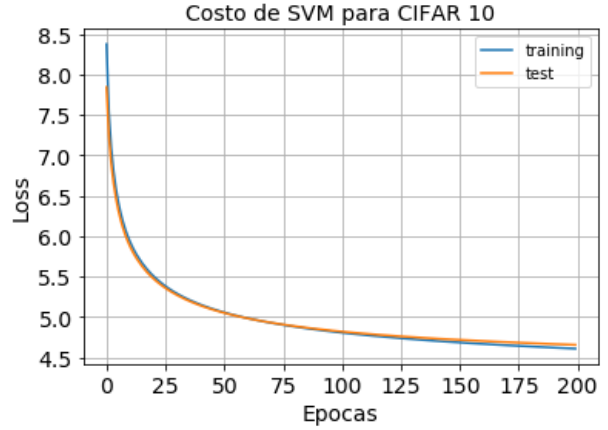


Figura 6: Costo de clasificación mediante el método SVM para el set de datos CIFAR-10.

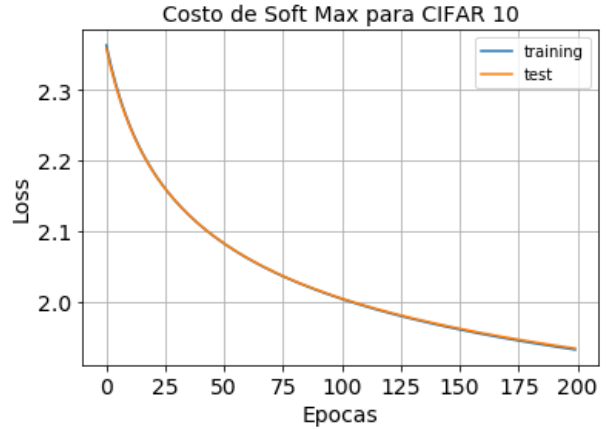


Figura 7: Costo de clasificación mediante el método SfM para el set de datos CIFAR-10.

En las figuras 8 y 9 se muestra la exactitud de haber realizado la clasificación mediante *SVM* y *SfM* respectivamente. Se puede observar que para aproximadamente 60 épocas *SVM* alcanza una exactitud de clasificación

de 35 % mientras que en las 200 épocas realizadas *SfM* no llega a alcanzar dicho valor.

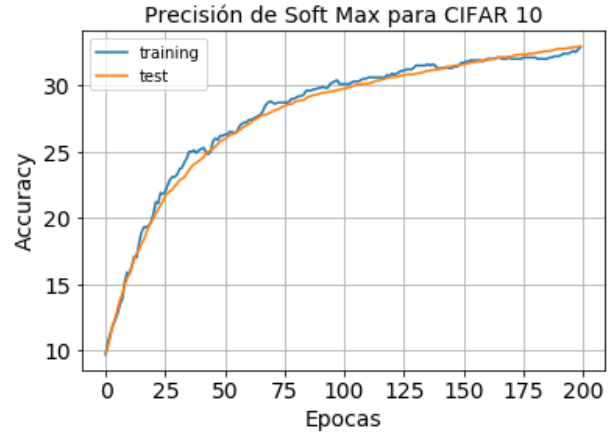


Figura 9: Exactitud de clasificación mediante el método *SfM* para el set de datos CIFAR-10.

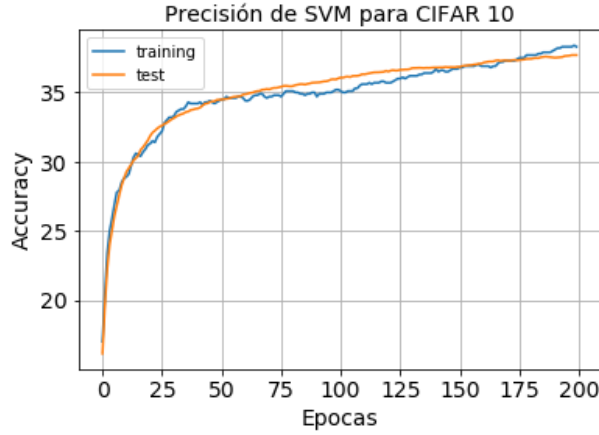


Figura 8: Exactitud de clasificación mediante el método *SVM* para el set de datos CIFAR-10.

Como conclusión *SVM* es más costoso, pero tiene una mayor exactitud, caso contrario de *SfM* de que es menos costoso y su exactitud es menor.

## VI. CONCLUSIONES

Se implementaron métodos de aprendizajes supervisados y no supervisados en el presente informe. Logrando entender los mecanismos de cada uno de los métodos.

Se comparó la diferencia de clasificación en un método supervisado y no supervisado tanto para la set de datos de MNIST como de CIFAR 10. La exactitud de clasificación de aprendizaje no supervisado es menor que el de supervisado. Se comparó la diferencia entre *SVM* y *SfM* para cada set de datos y se vio que depende del tipo de problema, puesto de que varía el costo y la exactitud. Así que depende del problema cuál es el método más conveniente para utilizar.