

Examen parcial 1 IA

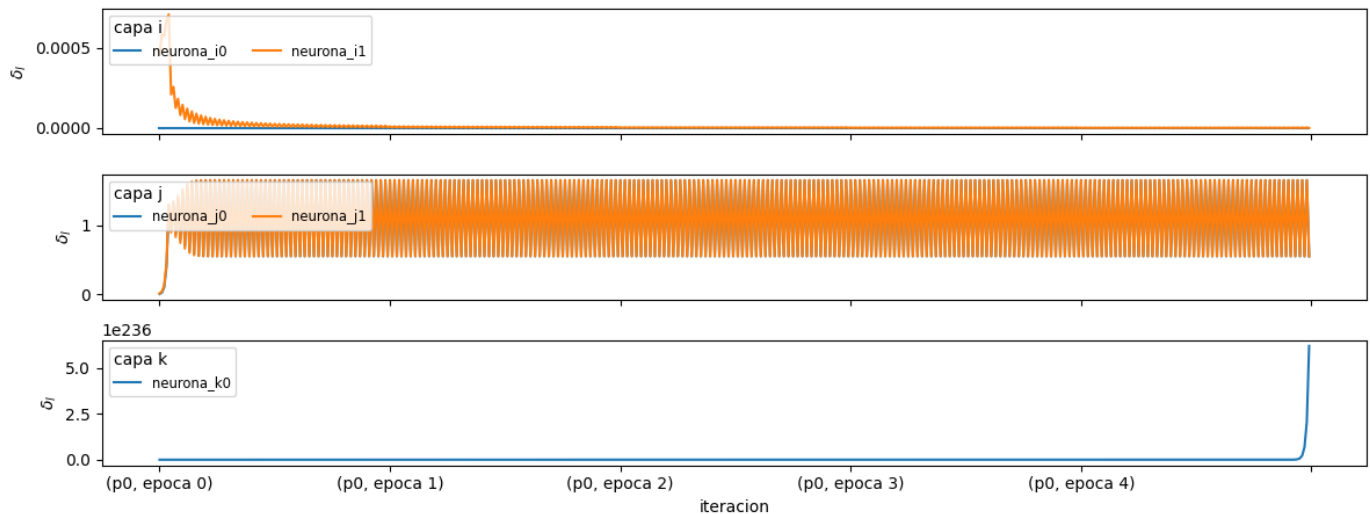
Ana Sofia Gutierrez

El modelo MLP tiene una configuracion de 3 capas:

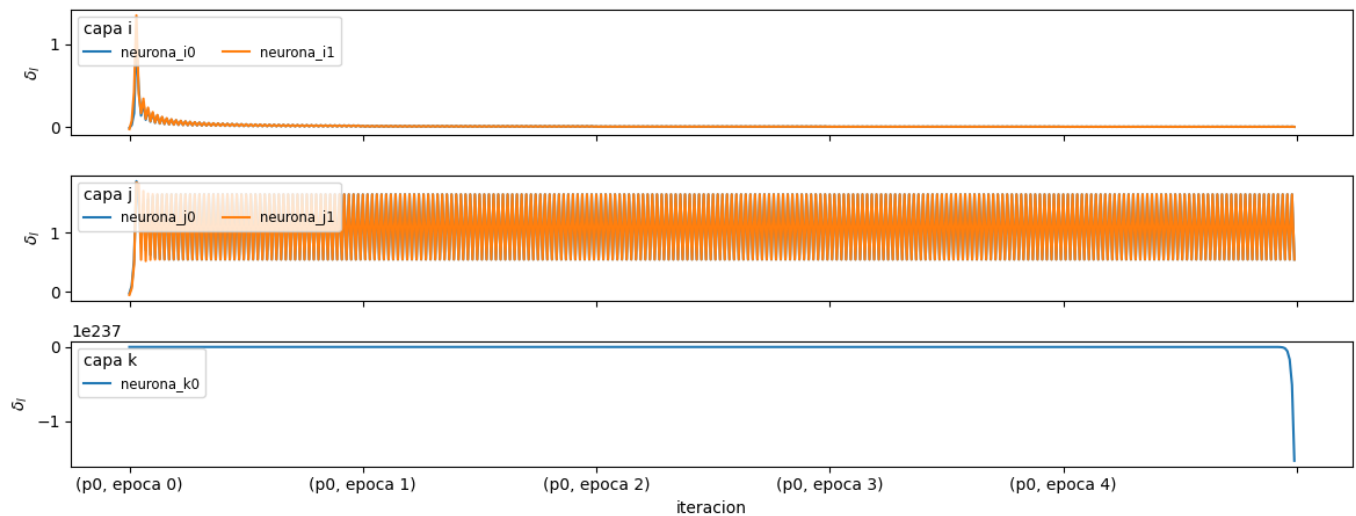
- Capa de entrada con 2 neuronas, funcion de activacion tanh
- Capa oculta con 2 neuronas, funcion de activacion sigmoide
- Capa de salida con 1 neurona, funcion de activacion lineal con $a=1$, $b=0$

Realizando descenso por el gradiente tomando los primeros 100 puntos del dataset, durante 5 epocas.

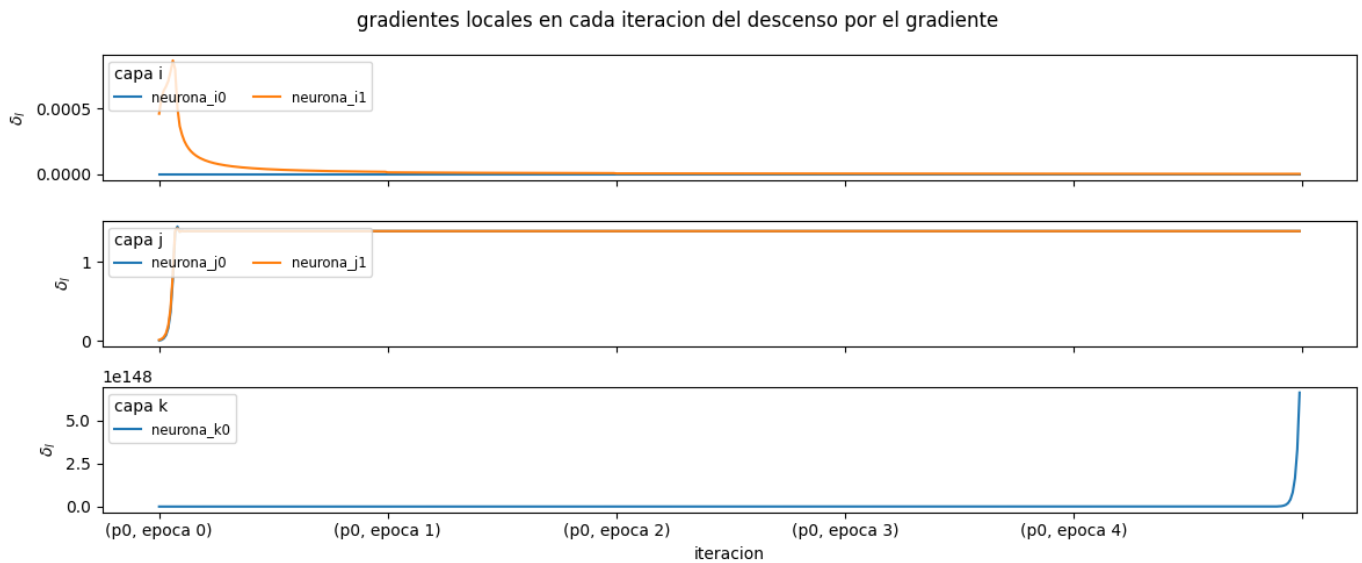
- Aprenda el dataset original en el modo secuencial usando como tasa de aprendizaje "eta" el valor 1
gradientes locales en cada iteracion del descenso por el gradiente



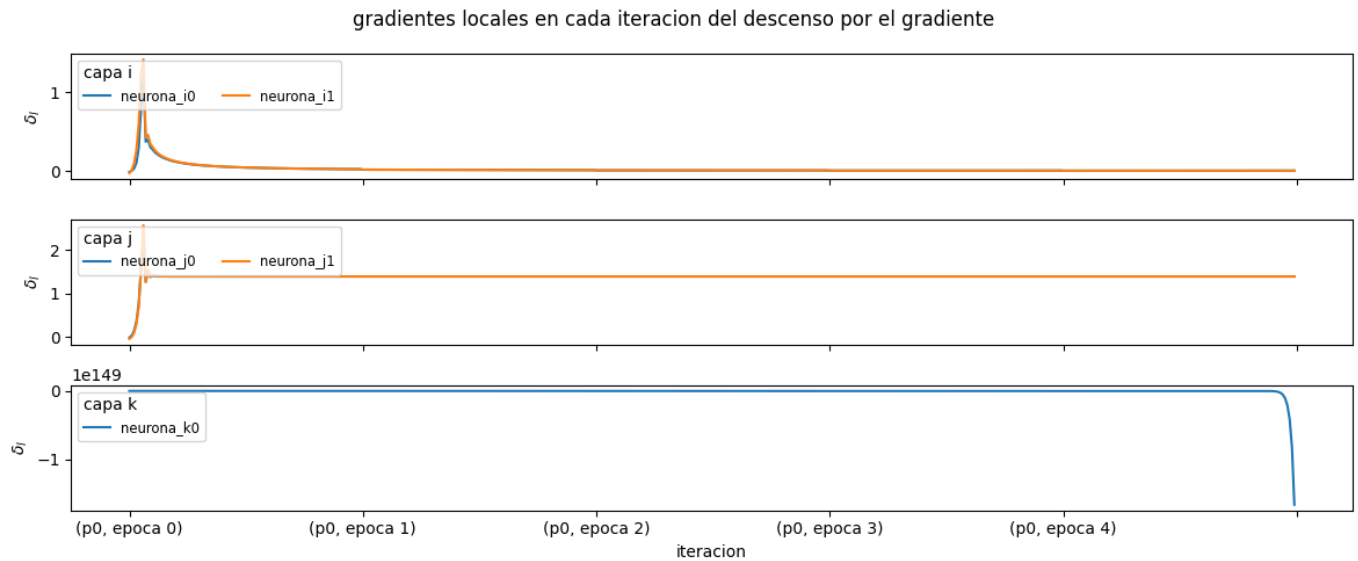
- Aprenda el dataset normalizado en el modo secuencial usando como tasa de aprendizaje "eta" el valor 1: realizando descenso por el gradiente tomando 100 puntos, durante 5 epocas
gradientes locales en cada iteracion del descenso por el gradiente



- Aprenda el dataset original con "eta" de 0.5

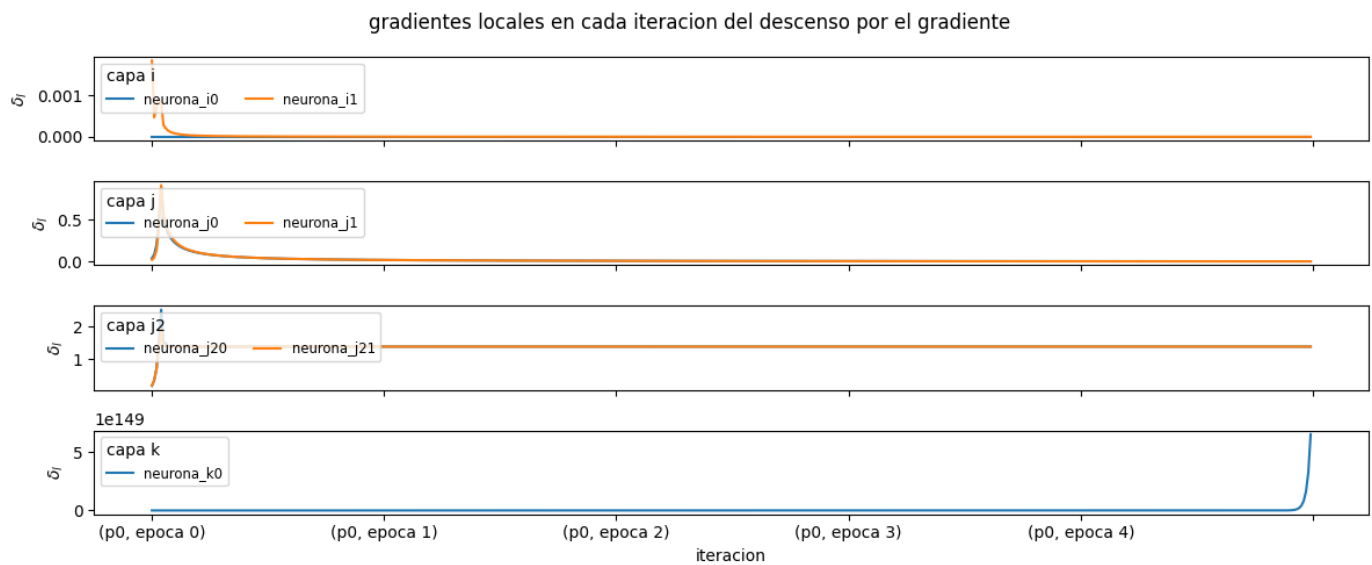


4. Aprende el dataset normalizado con "eta" de 0.5

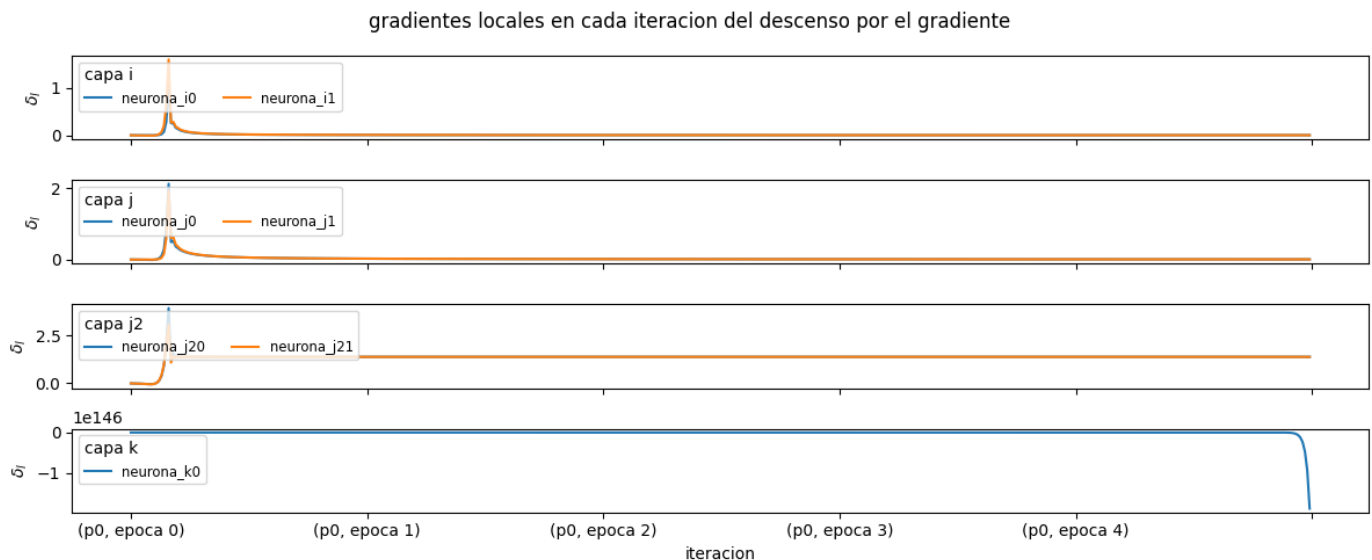


Segundo modelo MLP tiene una configuracion de 4 capas: Capa de entrada con 2 neuronas, funcion de activacion tanh. 2 capas ocultas con 2 neuronas, funcion de activacion sigmoide. capa de salida con 1 neurona, funcion de activacion lineal con $a=1$, $b=0$

5. Aprendizaje con $\eta=0.5$ sobre el dataset original



6. Con el dataset normalizado



Para un perceptron multicapa de 1 capa oculta con 2 neuronas en la capa oculta y otro de 2 capas ocultas con 2 neuronas por capa. **¿Cual aprende, que aprende y por que?**

Al comparar la informacion de los gradientes locales en cada capa, estamos mirando los factores que toman la informacion que va llegando del proceso backwards, para transmitirlos a la capa siguiente en el proceso. Esperaria que estos valores se vayan estabilizando en valores cercanos a cero conforme mejora en valor de la energia (error), el cual define el planteamiento de nuestro proceso de aprendizaje. Sin embargo, eso no sucede en ninguno de los modelos aca explorados. En ambos sucede que la capa de salida presenta un gradiente local muy grande en magnitud, que corresponde con que la energia del error tambien esta cambiando considerablemente con cada iteracion.

La normalizacion acota los valores de entrada al modelo en propagacion, y uno de los terminos del error en la retropropagacion. Pero al ser el proceso basado en multiplicaciones determinadas por la regla de la

cadena, esperaria que sea muy sensible a que haya valores altos. Sin embargo, la funcion elegida para las capas ocultas esta acotada por 1, incluso cuando el gradiente de la capa de salida entra a la oculta con valores muy grandes, va a ser multiplicado por valores menores a 1, que hacen que disminuya. Pensaria que este proceso no es muy estable, ya que en los primeros aprendizajes se presentan incluso oscilaciones en los gradientes locales de la capa oculta.

Finalmente, creo que ninguno de los modelos MLP planteados estan cumpliendo con el proceso ni de aprendizaje ni de entrenamiento. Los valores de los gradientes no estan “convergiendo” a cero y mi interpretacion anterior me lleva a pensar que hay mucha influencia de los valores iniciales de W , de los errores que van entrando al proceso y del orden en que son evaluados. Es cierto que los valores en la capa de entrada si van a cero, pero no creo que sea porque se esta cumpliendo el objetivo en general. La multiplicacion de los valores de los gradientes hace que rapidamente vayan a cero los valores gradiente en las ultimas capas de la retropropagacion, ya que en las capas ocultas se estan acotando los grandes gradientes que aparecen en la de salida. Y observar valores tan grandes en la capa de salida, que esta directamente determinada por el error en cada patron, me dice que la funcion de energia del error con la que estamos trabajando probablemente esta creciendo tambien.