

Teoría de Aprendizaje Estadístico y Teoría de Información: aplicaciones a Deep Learning

Trabajo Práctico

El trabajo práctico consiste en implementar algunos de los algoritmos vistos en clases. Los puntos pedidos a continuación son criterios de mínima, pudiendo ampliar el análisis a gusto del estudiante. Cada decisión a tomar o cada gráfico a mostrar debe estar debidamente explicado, en función de los conceptos de los fenómenos involucrados.

1. Tomar la base de datos MNIST y construir aleatoriamente un subconjunto de entrenamiento de 2000 muestras y dos subconjuntos de validación (distintos) de 200 muestras cada uno. El conjunto de testeo dejarlo en 10000 muestras. Inicialice las matrices de pesos con alguno de los criterios vistos en clase.
2. Construir un clasificador de 2 layers ocultos feed-forward (full connected) de 512 y 256 unidades respectivamente. Utilice uno de los conjuntos de validación para implementar un criterio de “early stop” y proponga un batchsize de 100. Grafique la cross-entropy y el porcentaje de clasificaciones correctas, tanto con el set de entrenamiento como con el de testeo. Reportar el error final de cada uno de los sets.
3. Implementar un “Denoising AutoEncoder” (DAE) con las siguientes características. La arquitectura será $784 - 512 - 256 - 512 - 784$ feed-forward con error cuadrático medio como función costo. Aplique ruido de rotación uniforme entre $(-\frac{\pi}{4}, \frac{\pi}{4})$ solamente en la entrada y efectúe un entrenamiento conjunto de todos los layers. Mostrar con imágenes como el DAE corrige las rotaciones (para imágenes de testeo) y graficar la función costo tanto para el set de entrenamiento como para el de testeo.
4. Repetir el punto 2 inicializando con los parámetros de los primeros layers ocultos del DAE (sin agregar el ruido).
5. Utilizar el segundo set de validación para fijar el batchsize. Grafique las cross-entropies y los porcentajes de clasificaciones correctas y extraiga conclusiones.

Hints

Se recomienda:

1. Utilizar el optimizador ADAM con un “learning rate” de 0,001.
2. Puede utilizar la función “tf.contrib.image.rotate” para rotar las imágenes.
3. En el caso del DAE se recomienda entrenar durante 500 epochs. Tenga en cuenta que los algoritmos no supervisados suelen tardar más en entrenarse que los supervisados.
4. En el punto 5 proponga como batchsize los valores 25, 50, 100, 200 y 400.