

Resultados de Examen final de Aprendizaje Profundo para MIGUEL LERDO DE TEJADA FLORES

ⓘ Las respuestas correctas ya no están disponibles.

Puntaje para este examen: 19 de 20

Entregado el 11 de jun en 9:27

Este intento tuvo una duración de 71 minutos.

Entregas de exámenes

[Intento 1: 19](#)

MIGUEL LERDO DE TEJADA FLORES no tiene más intentos
[← Volver a el examen](#)

Pregunta 11 / 1 pts

El modelo "Transformers" usa "positional encoding" para

- ☐ Compensar por la falta de orden secuencial de los tokens de entrada.
- ☐ Ocultar palabras clave en la secuencia de entrada.
- ☐ Simular pesos de atención entre los tokens de entrada.
- ☐ Compensar por la falta de celdas de atención, recuerdo y olvido.

Pregunta 21 / 1 pts

Una GAN condicional CGAN inyecta estructura al ruido Gaussiano que entra al generador. Esto se logra mediante:

- ☐ Modificar los parámetros de la distribución que entra al generador.
- ☐ Forzar a que el ruido de entrada tenga media = 0, y varianza = 1.
- ☐ Conectar la entrada del discriminador con la salida del generador.
- ☒ Incluir una lista de etiquetas en la entrada del generador.

Pregunta 31 / 1 pts

Una manera de verificar la distribución que tienen las distintas dimensiones de una representación latente de alta dimensionalidad (ej, 128-D) es:

- ☐ Usar varios datos de ejemplos para medir la mediana y la moda.
- ☐ Usar varios datos de ejemplos para medir la media y varianza.
- ☐ Usar varios datos de ejemplos para hacer una gráfica de nube de puntos.
- ☒ Usar varios datos de ejemplos para hacer una gráfica de cajas y brazos.

Pregunta 41 / 1 pts

Para una CNN, cuya primera capa contiene 8 filtros convolucionales de 5*5 píxeles, y recibe como entrada una imagen RGB de 220*220 píxeles, ¿Cuántos pesos existen en esa primera capa?

- ☐ 8*5*5 + 8
- ☐ 220*5*5 + 200*8
- ☐ 5*5*3 + 8
- ☒ 8*5*5*3 + 8

Pregunta 51 / 1 pts

¿Cuál de las siguientes no es una razón para el éxito de los modelos de aprendizaje profundo?

- ☐ La cascada profunda de no linealidades.
- ☒ El uso de combinaciones lineales.
- ☐ Hardware dedicado.
- ☐ Espíritu de trabajo colaborativo.

Pregunta 61 / 1 pts

Para un lote de entrenamiento de tamaño 64, ¿Cuál de las siguientes afirmaciones es correcta durante el proceso de entrenamiento?

- ☐ Se produce 1 solo error a la salida, y 1 solo gradiente que se retropropaga.
- ☒ Se producen 64 errores a la salida, y 1 solo gradiente que se retropropaga.
- ☐ Se producen 64 errores a la salida, y 64 gradientes que se retropropagan.
- ☐ Se produce 1 error a la salida, y 64 gradientes que se retropropagan.

Incorrecto

Pregunta 70 / 1 pts

Si ya tenemos el set de validación, el cual no es usado durante el entrenamiento y nos permite calcular el desempeño del modelo después de su entrenamiento, ¿por qué necesitamos un tercer set (prueba)?

- ☒ El set de validación de hecho sí se usa para aprender parámetros.
- ☐ El set de validación puede provenir de otra distribución de probabilidades generadora.
- ☐ El set de validación impacta la toma de decisiones del modelo.
- ☐ Para aumentar el soporte estadístico del desempeño evaluado.

Pregunta 81 / 1 pts

Una de las razones para preferir la función ReLU en las capas intermedias de una red es que:

- ☒ Disminuye la probabilidad de tener desvanecimiento de gradiente.
- ☐ Siempre produce errores positivos.
- ☐ Tienen un alto grado de soporte estadístico.
- ☐ Es diferenciable en todos lados.

Pregunta 91 / 1 pts

¿Cuáles son los dos parámetros estimados (no aprendidos) al usar el regularizados batch-normalization?

- ☐ Media y tasa de aprendizaje.
- ☐ Valor máximo y rango.
- ☐ Ninguna de las opciones.
- ☒ Media y desviación estándar.

Pregunta 101 / 1 pts

¿Cuántos parámetros hay en la siguiente red neuronal tipo MLP?

```
mip = Sequential([
    Input(shape=(30)),
    Dense(512, activation='relu'),
    Dense(128, activation='relu'),
    Dense(32, activation='relu'),
    Dense(6, activation='softmax')
])
```

- ☒ 85,862
- ☐ 85,184
- ☐ 72,385
- ☐ 85,697

Pregunta 111 / 1 pts

Una manera de obtener vectores raros (sparse) en alguna capa de una red neuronal es:

- ☐ Usando un regulador L2 en esa capa.
- ☒ Usando un regulador L1 en esa capa.
- ☐ Usando la divergencia Kullback-Leibler en esa capa.
- ☐ Haciendo la capa tan compacta como sea posible.

Pregunta 121 / 1 pts

¿Sin considerar sesgos, en una vanilla RNN, cuál es el número de pesos de recurrencia en una capa de recurrencia con 9 perceptrones?

- ☒ 81
- ☐ 18
- ☐ 90
- ☐ 9

Pregunta 131 / 1 pts

Una función de salida utilizada para los problemas de regresión multivariada es:

- ☐ Cross entropy
- ☐ Divergencia Kullback-Leibler
- ☒ Mean absolute error
- ☐ ReLU

Pregunta 141 / 1 pts

¿Cuál es el set de datos con el que dejamos que el modelo ajuste sus pesos de manera automática?

- ☐ Prueba.
- ☐ Validación.
- ☒ Entrenamiento.
- ☐ Cross-validation.

Pregunta 151 / 1 pts

La detección de anomalías mediante el uso de autoencoders supone que:

- ☐ El error de reconstrucción de las anomalías es notablemente menor que el de los datos normales.
- ☒ El error de reconstrucción de las anomalías es notablemente mayor que el de los datos normales.
- ☐ El espacio latente de las anomalías es totalmente distinto al espacio latente de los datos normales.
- ☐ El espacio latente de las anomalías es denso, mientras que el de los datos normales es raro.

Pregunta 161 / 1 pts

¿Cuál función de activación es más adecuada para, la capa de salida, de una red convolucional que realiza reconstrucción de imágenes, donde los valores de los píxeles son números reales entre 0 y 1?

- ☐ Softmax.
- ☒ Sigmoide.
- ☐ Leaky ReLU.
- ☐ Tanh.

Pregunta 171 / 1 pts

Con el método de retropropagación (backpropagation):

- ☐ Los pesos más cercanos al centro de la red (representación latente) sufren actualizaciones de mayor magnitud.
- ☐ Los pesos más cercanos al inicio de la red sufren actualizaciones de mayor magnitud.
- ☒ Los pesos más cercanos al final de la red sufren actualizaciones de mayor magnitud.
- ☐ Los pesos con regularización sufren actualizaciones de mayor magnitud.

Pregunta 181 / 1 pts

Es una suposición que se hace comúnmente sobre los datos de entrenamiento y validación:

- ☒ Son generados por la misma distribución de probabilidades.
- ☐ No tienen sobre ajuste.
- ☐ Sin fáciles de conseguir pero difíciles de etiquetar.
- ☐ Son generados por la misma persona.

Pregunta 191 / 1 pts

La diferencia principal entre un autoencoder vanilla (AE) y uno variacional (VAE), es que:

- ☒ Cada variable del espacio latente de un VAE se aproxima a una distribución normal, mientras que la del AE no necesariamente lo hace.
- ☐ El espacio latente del VAE siempre es raro, mientras que el del AE no tiene garantía de serlo.
- ☐ El espacio latente del VAE tiene más elementos que el del AE.
- ☐ Cada variable del espacio latente de un AE se aproxima a una distribución normal, mientras que la del VAE no necesariamente lo hace.

Pregunta 201 / 1 pts

¿Cuál es una de las ventajas de hacer descenso de gradiente por lotes?

- ☐ Es mejor que el hiperplano considere la convexidad del set de entrenamiento.
- ☐ Se evita que haya desvanecimiento del gradiente.
- ☐ Es mejor que el hiperplano se ajuste, de manera individual, a cada uno de los puntos del set de entrenamiento.
- ☒ Es mejor que el hiperplano se ajuste, al mismo tiempo, a una muestra representativa de todos los datos de entrenamiento.

Puntaje del examen: 19 de 20