

Pregunta 1

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 2

1 / 1 pts

- ☐ 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.

☐ 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 4 1 / 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

¿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 6 1 / 1 pts

☐ 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.

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.

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 9 1 / 1 pts

☐ Media y tasa de aprendizaje.

☐ Valor máximo y rango.

☐ Ninguna de las opciones.

☒ Media y desviación estándar.

```
Dense(128, activation='relu'),
Dense(32, activation='relu'),
Dense(6, activation='softmax')
})
```

Pregunta 11

1 / 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 12

1 / 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

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

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

Pregunta 14 1 / 1 pts

pesos de manera automática?

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

Pregunta 15 1 / 1 pts

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

Pregunta 16

1 / 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.

Con el método de retropropagación (backpropagation):

- 1. Los pesos más cercanos al centro de la red (representación latente) sufren actualizaciones de mayor magnitud.
- 2. Los pesos más cercanos al inicio de la red sufren actualizaciones de mayor magnitud.
- 3. Los pesos más cercanos al final de la red sufren actualizaciones de mayor magnitud.

Pregunta 18

1 / 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.

Pregunta 19 1 / 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.

Cada variable del espacio latente de un AE se aproxima a una distribución normal, mientras que la del VAE no necesariamente lo hace.

Pregunta 20 1 / 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.