

1. Verdadero o falso: el objetivo de la destilación del conocimiento es optimizar la implementación de la red:

1 / 1 punto

- ☒ Falso
- ☐ Verdadero

✓ **Correcto**

¡Exactamente! En lugar de optimizar, la destilación busca crear un modelo más eficiente.

2. En la destilación del conocimiento, el maestro será capacitado utilizando un _____.

1 / 1 punto

- ☐ GoogleLeNet
- ☐ Un objetivo blando
- ☒ Una función objetivo estándar
- ☐ divergencia KL

✓ **Correcto**

¡Lo superó! Esto busca maximizar la precisión del modelo.

3. Verdadero o falso: DistilBERT es una versión más grande de BERT con una arquitectura modificada, pero con el mismo número de capas.

1 / 1 punto

- ☐ Verdadero
- ☒ Falso

✓ **Correcto**

¡Estás bien! es una versión más pequeña de BERT: redujeron la cantidad de capas y mantuvieron el resto de la arquitectura idéntica

4. Verdadero o falso: en la destilación del conocimiento, la red de "maestros" se implementa en la producción, ya que es capaz de imitar las complejas relaciones de funciones de la red de "estudiantes".

1 / 1 punto

☐ Verdadero

☒ Falso

✓ **Correcto**

¡Exactamente! En realidad, es la red de "estudiantes" la que se implementó para imitar la red de "maestros".

5. Para un problema de clasificación de clases múltiples, ¿cuáles de las siguientes afirmaciones son verdaderas con respecto a las funciones de costo de capacitación de las redes de "estudiante" y "maestro"? (Seleccione todas las que correspondan)

1 / 1 punto

☒ Los objetivos blandos codifican más información sobre el conocimiento aprendido por el profesor que su predicción de clase de salida por ejemplo.

✓ **Correcto**

¡Así es! Los objetivos suaves proporcionan más información que la clase de salida predicha por ejemplo, ya que incluyen información sobre todas las clases por ejemplo de entrenamiento a través de la distribución de probabilidad.

☐ La red de maestros está capacitada para maximizar su precisión y la red de estudiantes utiliza una función de costo para generar las mismas clases que la red de maestros.

☐ Ambos comparten las mismas funciones de costo,

☒ La red de maestros está entrenada para maximizar su precisión y la red de estudiantes usa una función de costo para aproximar las distribuciones de probabilidad de las predicciones de la red de maestros.

✓ **Correcto**

¡Así es!

6. Cuando la temperatura softmax ____, los objetivos suaves definidos por la red de maestros se vuelven menos informativos

1 / 1 punto

☐ es igual a 1

☐ aumenta

☒ disminuye

**Correcto**

¡Así es! La suavidad de la distribución del maestro es peor, por lo tanto, menos informativa.

7. Generalmente, la destilación del conocimiento se realiza combinando dos funciones de pérdida e involucra varios hiperparámetros. Aquí, L_h es la pérdida de entropía cruzada de las etiquetas duras y LKL es la pérdida de divergencia de Kullback-Leibler de las etiquetas maestras. ¿Cuáles de las siguientes afirmaciones son correctas sobre los hiperparámetros de la destilación del conocimiento? (Seleccione todas las que correspondan)

1 / 1 punto

- ☐ Cuando calculamos la pérdida "estándar" entre las probabilidades de clase predichas del estudiante y las etiquetas "duras" de la verdad básica, usamos el mismo valor de la temperatura softmax T para calcular el softmax en los logits del maestro.
- ☒ En caso de un gran aumento de datos después de entrenar la red del maestro, el hiperparámetro alfa debe ser alto en la función de pérdida de la red del estudiante

**Correcto**

¡Eso es correcto! Este parámetro alfa alto reduciría la influencia de las etiquetas duras que sufrieron perturbaciones agresivas debido al aumento de datos.

- ☐ En caso de un gran aumento de datos después de entrenar la red del maestro, el hiperparámetro alfa debe ser bajo en la función de pérdida de la red del estudiante
- ☒ Cuando calculamos la pérdida "estándar" entre las probabilidades de clase predichas del estudiante y las etiquetas "duras" de la verdad básica, usamos un valor de la temperatura softmax T igual a 1

**Correcto**

¡Así es! De esta forma, la función de pérdida de estudiantes sería una función softmax clásica.