

Seminario Khipu 2023

Control Final

Nombre:

C.I.:

- Posgrado
- Actualización
- Extracurricular

El control es sin material y tiene una duración de 1h15.

ML Fundamentals [Luciana Ferrer]

1. Indique si las siguientes afirmaciones son verdaderas o falsas
 - a. Cuando la matriz de costo tiene unos en todas las entradas y ceros en la diagonal, el expected cost coincide con 1-accuracy **V**
 - b. La cross-entropy es una medida más exhaustiva de la performance que el accuracy. **V**
2. Seleccionar todas las opciones correctas.
Cuando una clase minoritaria es más importante que otras (por ejemplo en aplicaciones médicas):
 - a. El accuracy es una medida de performance razonable en este escenario. **F**
 - b. La medida F_β es mejor que el accuracy en estos casos. **V**
 - c. Puede haber clasificadores con valores relativamente altos de F_β , pero que no son útiles porque tienen valores de Normalized EC mayores que 1. **V**

RL 1 [Pablo Castro]

3. Marque las opciones correctas
 - a. Reinforcement learning es típicamente formalizado utilizando MDP (Markov Decision Process). **V**
 - b. En RL la política óptima es conocida y el algoritmo debe aprender las posibles acciones con las que cuenta el agente. **F**
 - c. La función de valor $V(s)$ no depende de la recompensa inmediata obtenida en el estado s al realizar una acción a ($R(s,a)$), sino de la recompensa futura. **F**

4. Considerando el algoritmo de Policy Iteration utilizado en RL, ¿cuál de las siguientes formulaciones de la ecuación de Bellman es correcta?

- a. $Q^*(s) = R(s) + \gamma \sum_{s' \in S} P(s, a)(s')Q^*(s')$ **F**
- b. $Q(s, a) = R(s, a) + \gamma \sum_{s' \in S} P(s, a)(s')Q(s', \pi(s'))$ **Correcta**
- c. $Q(s, a) = R(s, a) + \sum_{s' \in S} P(s, a)(s')Q(s', \pi(s'))$ **F, falta gamma**
- d. $Q(s, a) = R(s, a) + \gamma \sum_{s' \in S} P(s, a)(s')Q(s', a)$ **F falta pi**

RL 2 [Doina Precup]

5. Cuales de las siguientes afirmaciones son ciertas sobre las diferencias entre Temporal Difference Learning y Monte Carlo.

- a. Monte Carlo actualiza los valores de los estados en cada paso de tiempo, mientras que TDL lo hace solo al final del episodio. **F**
 - b. Montecarlo utiliza todos los pasos de un recorrido para actualizar los valores de los estados, mientras que TD utiliza solo el siguiente estado para actualizar los valores. **V**
 - c. Cuanto más pasos se utilizan en el algoritmo n-step TD menor es la varianza y menor el sesgo (bias) de los valores obtenidos. **F**
6. ¿Cuál de las siguientes afirmaciones es verdadera sobre el uso de aproximación de funciones para resolver el problema de dimensionalidad (Curse of dimensionality) en el aprendizaje por refuerzo?
- a. La aproximación de funciones no es efectiva para resolver el problema de dimensionalidad en el aprendizaje por refuerzo. **F**
 - b. La aproximación de funciones es una técnica comúnmente utilizada para resolver el problema de dimensionalidad en el aprendizaje por refuerzo. **V**
 - c. La aproximación de funciones sólo se aplica a problemas de baja dimensionalidad en el aprendizaje por refuerzo. **F**

Generative Models [Ruben Villegas]

7. ¿Cuál de las siguientes oraciones define mejor el funcionamiento de los Transformers en inferencia?
- Le paso como entrada una oración al encoder y sólo con eso el decoder me devuelve una oración del mismo largo a la salida con la traducción.
 - Le paso como entrada una oración al encoder, junto con un vector de ruido a la entrada del decoder para que este genere una oración entera a la salida con la traducción correspondiente.
 - Le paso como entrada una oración al encoder, junto con un token de inicio a la entrada del decoder para que este genere una oración entera a la salida con la traducción correspondiente.
 - Le paso como entrada una oración al encoder, junto con un token de inicio a la entrada del decoder, y realimento a éste con las predicciones de sí mismo hasta tener una oración completa. (Correcta)**

Computer Vision [Jorge Sánchez]

8. Indique si las siguientes afirmaciones son verdaderas o falsas:
- Las redes convolucionales son equivariantes a translaciones **V**
 - La equivariancia a translaciones en las redes convolucionales es una consecuencia del weight sharing de los kernels **V**
 - Las redes convolucionales son equivariantes a rotaciones **F**
 - Las capas de pooling introducen no-linealidad pero reducen el receptive field **F**
9. Indique si las siguientes afirmaciones son verdaderas o falsas:
- Los transformers tienen la capacidad de incluir información de contexto global, gracias al mecanismo de self attention **V**
 - Un transformer sin positional encoding es equivariante ante permutaciones en la entrada. Agregando un positional encoding, es posible hacerlo invariante a estas permutaciones. **F**
 - El mecanismo de multi-head attention consiste en una combinación lineal de las columnas de V , ponderadas por pesos que dependen no linealmente del producto escalar entre columnas de Q y de K . **V**

GNN I y II [Alejandro Ribeiro y Gonzalo Mateos]

10. Así como sucede con las CNNs, una red neuronal completamente conectada (FCNN) es también una generalización de una GNN, por lo que teóricamente debería ser capaz de obtener mejores resultados en entrenamiento. Marque hasta dos propiedades que explican porqué, a pesar de esto, la GNN generaliza mejor que una FCNN en escenarios no vistos durante el entrenamiento:
- a. La salida de una GNN es invariante a permutaciones. Es decir, un re-etiquetado de los nodos no cambia la salida. **F**
 - b. La salida de una GNN es equivariante a permutaciones. Es decir, un re-etiquetado de los nodos cambia la salida de la misma forma. **V**
 - c. Una GNN, una vez entrenada, se puede usar en un grafo de cualquier tamaño, incluso en grafos de mayor tamaño a los vistos durante el entrenamiento. Esto es imposible para una FCNN. **V**
 - d. Pequeños cambios en el grafo producen grandes cambios a la salida de la GNN, lo que le permite ser muy discriminador. **F**
11. Suponga que cuenta con una red de citas entre papers (los nodos son los papers, y hay una arista entre dos nodos si uno cita al otro). Cuenta además con una bolsa de N palabras y con un vector $\{0, 1\}^N$ por paper indicando si la palabra está en él o no. El objetivo es entrenar una GNN para predecir a cuál de 7 áreas pertenece cada paper (biología, matemática, etc.), para lo cual cuenta además con una etiqueta indicando el área para varios de los papers en su dataset. Indique cuáles de estas afirmaciones son verdaderas:
- a. Se trata de un problema de *graph classification*, donde queremos categorizar el grafo en una de las 7 áreas posibles a partir de la señal de entrada en cada nodo. **F**
 - b. Se trata de un problema de *node classification*, donde queremos categorizar cada nodo en una de las 7 áreas posibles a partir de la señal de entrada en cada nodo. **V**
 - c. Se trata de un problema de link prediction, donde queremos predecir a qué área (representadas como otros siete nodos) está conectado cada paper. **F**

LLM [Nando de Freitas]

12. En la arquitectura estándar de Transformers. ¿Cuáles de las siguientes combinaciones de técnicas se utilizan específicamente para acelerar el entrenamiento?
- a. Multi-Head Attention, Masking, Layer Normalization
 - b. Masking, Layer Normalization, Residual Connections (Correcta)**
 - c. Residual Connections, Attention, Masking
 - d. Self-Attention, Layer Normalization, Residual Connections.

Ethics and Fairness [Sara Hooker]

13. ¿Cuáles de las siguientes afirmaciones son verdaderas?

- a. La justicia en un sistema de Aprendizaje Automático se define como las preferencias acerca de cómo debería comportarse el sistema respecto a un sub-conjunto de características de los datos. O sea, es un concepto intrínsecamente subjetivo y puede cambiar tanto en el tiempo como en el espacio. **V**
- b. Una técnica de compresión de algoritmos de aprendizaje profundo es eliminar varios pesos y hacer más esparza las operaciones. Esto trae naturalmente como consecuencia una pérdida de precisión, que si la compresión está bien diseñada debería ser mínima. Si esta pérdida de precisión es pequeña, entonces el impacto en la justicia del sistema también es pequeño. **F**
- c. Tomar en cuenta únicamente la precisión como desempeño del sistema hace que éste reproduzca clichés que se encuentran en la base de datos de entrenamiento. Afortunadamente, es sencillo anotar la base de datos para poder tomar en cuenta estos sesgos a la hora de entrenar. **F**