



Repaso Control Largo 2

1. Árboles de Decisión y Random Forest

1. ¿Es recomendado hacer un árbol de decisión con la mayor profundidad posible?
 - a) Sí, ya que nos permite maximizar el rendimiento del modelo en el set de entrenamiento.
 - b) No, ya que no necesariamente las características que tenemos son predictivas para características que no hayamos visto con anterioridad.
 - c) Sí, ya que está demostrado que los árboles generalizan mejor a medida que son más profundos.
 - d) No, ya que puede significar una sobrecarga en la capacidad de cómputo de nuestro computador.
2. ¿Cuál de las siguientes afirmaciones sobre un Random Forest como un método de clasificación es correcta? (Puedes seleccionar múltiples alternativas).
 - a) Conviene utilizar la mayor cantidad de atributos para cada árbol, ya que así maximizamos la correlación entre los árboles.
 - b) Se podría ver a un Random Forest como un sistema democrático, donde cada árbol tiene un voto a la hora de determinar la clase.
 - c) Conviene tener una cantidad restringida de atributos a elegir para cada nodo en cada árbol, ya que así minimizamos la correlación entre los árboles.
 - d) Suele clasificar mejor que un árbol de decisión singular.
3. Existen diversos motivos por el cual utilizamos múltiples clasificadores débiles a la hora de armar un Random Forest. De las siguientes afirmaciones, ¿cuál NO es un motivo para ello?
 - a) Para reducir la correlación entre todos los árboles del bosque.
 - b) Para poder entrenar la mayor cantidad de modelos posibles utilizando subconjuntos del set de entrenamiento.
 - c) Para evitar el subajuste del modelo.
 - d) Para poder entrenar la mayor cantidad de modelos posibles utilizando subconjuntos de las características totales.

2. Support Vector Machine

4. ¿Cuáles de las siguientes aseveraciones es correcta con respecto a la relación entre el margen y $|\vec{w}|$?
- a) El margen es inversamente proporcional a $|\vec{w}|$.
 - b) El margen es directamente proporcional a $|\vec{w}|$.
 - c) El ancho total del margen es igual a $\frac{1}{|\vec{w}|}$.
 - d) El ancho total del margen es igual a $|\vec{w}|$.
5. ¿Qué utilidad entrega el utilizar constantes C independientes para cada clase en un SVM?
- a) Permite penalizar los ejemplos duplicados de cada clase.
 - b) Permite manejar el desbalance de las clases.
 - c) Permite descomponer el problema de optimización para una resolución más eficiente.
 - d) Permite evitar el uso de los multiplicadores de Lagrange en la resolución.
6. ¿Cuáles de las siguientes aseveraciones sobre un kernel-SVM son correctas? (puedes seleccionar múltiples alternativas)
- a) Su superficie de decisión en el espacio de características transformado no es necesariamente lineal.
 - b) Su superficie de decisión en el espacio de características original no es necesariamente lineal.
 - c) Tienen más riesgo de sobreajuste.
 - d) El kernel calcula la transformación de los vectores de características al nuevo espacio.
7. *Entrenar un SVM sobre todo un conjunto de entrenamiento es lo mismo que entrenarlo sólo en los vectores de soporte de ese mismo conjunto.* Respecto a la afirmación anterior podemos afirmar que:
- a) La afirmación es verdadera.
 - b) La afirmación es falsa.
8. Cuando trabajamos con SVM en muchas ocasiones podemos encontrarnos con problemas no linealmente separables. ¿Cómo podemos afrontar este problema para encontrar, de todos modos, un hiperplano separador? (Puedes seleccionar múltiples respuestas).
- a) Incluir variables auxiliares que actúen como una relajación del problema original y permita que algunos datos se encuentren en el lado erróneo de la superficie de decisión.
 - b) Aumentar los grados de libertad del hiperplano, a modo de que separe correctamente los datos a pesar de que el problema no sea separado por una recta.
 - c) Entrenar un SVM multiclase, donde se clasifiquen bajo otra categoría aquellos datos que se encuentran en el lado incorrecto de la superficie de decisión.
 - d) Transformar el espacio de características a modo de mapear los datos a un espacio en donde si son linealmente separables.

3. Redes Neuronales

9. El conjunto de datos HASYv2 contiene imágenes en blanco y negro de 369 íconos y símbolos matemáticos distintos, cada una con tamaño 30x30 píxeles. Algunas de ellas se muestran a continuación:



Figura 1: Fragmento del dataset HASYv2

Si quisiéramos entrenar una red neuronal para clasificar estas imágenes utilizando el valor de cada uno de sus píxeles como entrada, ¿cuántas neuronas necesitará tener la capa de entrada de la red?

- a) 30 neuronas
- b) 369 neuronas
- c) 900 neuronas
- d) 11070 neuronas

Utilice la figura 2 para responder las siguientes preguntas:

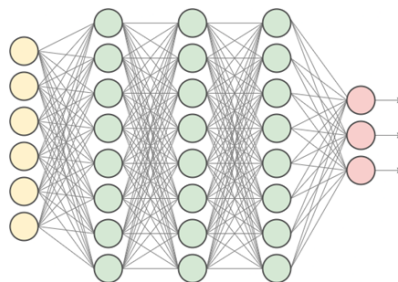


Figura 2: Red neuronal artificial de ejemplo

10. Respecto al número de neuronas de la red, podemos afirmar que posee:
- a) 6 neuronas en la capa de entrada, 3 en cada capa oculta y 3 en la capa de salida.
 - b) 6 neuronas en la capa de entrada, 8 en cada capa oculta y 3 en la capa de salida.
 - c) 3 neuronas en la capa de entrada, 3 en cada capa oculta y 6 en la capa de salida.
 - d) 3 neuronas en la capa de entrada, 8 en cada capa oculta y 6 en la capa de salida.

11. Con respecto a la entrada de la red, podemos afirmar que se debe tratar de:
- a) 6 vectores de cualquier largo.
 - b) Una matriz de 6x6.
 - c) Un vector de 6 elementos.
 - d) 6 imágenes de 28x28.
12. Con respecto a la salida de la red, podemos afirmar que se debe tratar de:
- a) Un vector de 3 elementos
 - b) Un vector de 6 elementos
 - c) Una matriz de 8x3
 - d) Un único valor numérico (0 o 1)
13. ¿Cómo propaga el algoritmo de backpropagation los pesos desde la capa de salida al resto de las capas?
- a) Se computa cuáles neuronas contribuyen más a la función de pérdida y se escogen nuevos valores para sus pesos de forma aleatoria, mejorando el desempeño en cada nueva iteración del algoritmo.
 - b) Utilizando regla de la cadena, el algoritmo computa la derivada parcial de la función de error con respecto al peso de la respectiva neurona.
 - c) Se promedian las derivadas parciales de la función de pérdida con respecto a cada uno de los pesos y luego se le resta este valor a cada uno de los pesos de las neuronas, moviéndolos en la dirección del gradiente.
 - d) Se escogen al azar un cierto número de neuronas de la última capa y se actualizan aleatoriamente sus pesos, luego, se repite este proceso por cada capa anterior con un menor número de neuronas dadas por un parámetro α .
14. ¿Con cuál(es) de los siguientes problemas puede trabajar un MLP de forma exitosa? (Puedes seleccionar múltiples alternativas)
- a) Regresión sobre datos lineales
 - b) Clasificación de datos linealmente separables
 - c) Regresión sobre datos no lineales
 - d) Clasificación de datos no linealmente separables

4. Aprendizaje Reforzado

15. ¿Cuál(es) de las siguientes opciones forma parte de un Proceso de Decisión de Markov? (Puedes seleccionar más de una alternativa)
- a) Un set de recompensas
 - b) Un set de acciones
 - c) Un set de probabilidades de transición entre estados
 - d) Un set de políticas
16. ¿Por qué utilizamos un factor de descuento en el cálculo del retorno G_t ?
- a) Para penalizar acciones que sean distintas a la política actual.
 - b) Para ajustar la importancia que el agente les asignará a recompensas más lejanas en el tiempo.
 - c) Para evitar cambiar la política del agente constantemente.
 - d) Para considerar en mayor o menor escala el próximo movimiento del oponente.
17. En el contexto de Q-Learning ¿En base a qué variables estima la función acción valor Q_π el valor del retorno? (Puedes seleccionar más de una alternativa)
- a) A los estados anteriores s_0 a $s_{(t-1)}$.
 - b) Al estado actual s_t .
 - c) Al valor esperado del estado anterior $G_{(t-1)}$.
 - d) A la acción tomada a_t .
18. ¿Cómo es computado el retorno esperado del agente en el siguiente estado?
- a) Se estima la recompensa esperada de ejecutar la acción a_t mediante algún algoritmo de aprendizaje.
 - b) Se suman todas las recompensas esperadas en un horizonte de tiempo h , ponderando cada una por un factor de descuento.
 - c) Se aprende mediante prueba y error (en forma de la recompensa), almacenándose en la función Q_π .
 - d) No se puede estimar el retorno esperado del agente.
19. ¿Cómo podemos definir a la política óptima de un agente?
- a) Aquella política que maximiza la recompensa a lo largo del tiempo.
 - b) Aquella política que minimiza la recompensa a lo largo del tiempo.
 - c) Aquella política que minimiza el número de acciones para resolver un problema.
 - d) Aquella política que minimiza la recompensa adversaria al jugar con un contrincante.

Pauta

1. Árboles de Decisión y Random Forest

1. ¿Es recomendado hacer un árbol de decisión con la mayor profundidad posible?
 - a) Incorrecto. Si bien una mayor profundidad puede significar un mejor rendimiento en entrenamiento, esto puede llevar a sobreajuste en el set y una mala generalización en la aplicación.
 - b) **Correcto. Las características elegidas para los niveles más profundos de un árbol tienden a sobreajustarse a la porción de datos de entrenamiento en que se basa, debido a la reducción progresiva de estos al aumentar la profundidad.**
 - c) Incorrecto. Un árbol no necesariamente generalizará mejor a mayor profundidad, esto debido a que algunos datos en profundidades más lejanas pueden corresponder a sobreajustes sobre el set de entrenamiento, por tanto, no es recomendable que el árbol sea lo más profundo posible.
 - d) Incorrecto. El algoritmo utilizado para crear un árbol de decisión es relativamente simple y no requiere gran capacidad de cómputo, luego, su evaluación es sencilla y solo se trata de una serie de comparaciones. Es correcto que el árbol no debe ser lo más profundo posible, pero no debido a este motivo.
2. ¿Cuál de las siguientes afirmaciones sobre un Random Forest como un método de clasificación es correcta? (Puedes seleccionar múltiples alternativas).
 - a) Incorrecto. Al maximizar la correlación entre los árboles estamos básicamente entrenando muchos árboles idénticos, todos propensos a caer en los mismos sesgos. Siempre trataremos de que nuestros árboles tengan la menor correlación posible.
 - b) **Correcto. Para la toma de decisión, cada árbol tiene un voto (que puede tener distintos pesos o no) para definir la categoría.**
 - c) **Correcto. Minimizar la correlación evita que múltiples árboles recaigan en los mismos sesgos.**
 - d) **Correcto. Efectivamente, al tratarse de un ensamble de árboles de decisión, suele tener mejor rendimiento que un modelo por su cuenta.**
3. Existen diversos motivos por el cual utilizamos múltiples clasificadores débiles a la hora de armar un Random Forest. De las siguientes afirmaciones, ¿cuál NO es un motivo para ello?
 - a) Incorrecto. Utilizar varios clasificadores débiles permite entrenar múltiples de estos con distintos subconjuntos del set total de datos.
 - b) Incorrecto. Modelos más débiles necesitan menos datos para entrenar, por lo que podemos entrenar más de ellos utilizando menos datos. Recuerda que los ensambles de modelos buscan reducir la correlación entre cada uno de ellos al mínimo posible.
 - c) **Correcto. Si bien el rendimiento del modelo puede diferir dependiendo del clasificador utilizado, el ensamble de modelos más simples en general conlleva un aumento en el rendimiento comparado a un simple árbol de decisión.**
 - d) Incorrecto. Modelos más débiles usan menos características para llevar a cabo su decisión. Recuerda que los ensambles de modelos buscan reducir la correlación entre cada uno de ellos al mínimo posible.

2. Support Vector Machine

4. ¿Cuáles de las siguientes aseveraciones es correcta con respecto a la relación entre el margen y $|\vec{w}|$?
- a) **Correcto. La formulación geométrica del SVM permite establecer esto por construcción.**
 - b) Incorrecto. La formulación geométrica del SVM establece una condición distinta a esta.
 - c) Incorrecto. Esta cantidad corresponde a una proporción del margen, no a su totalidad.
 - d) Incorrecto. La formulación geométrica del SVM no establece una condición de proporcionalidad con respecto a $|\vec{w}|$, por lo que el ancho no puede ser igual a esa cantidad.
5. ¿Qué utilidad entrega el utilizar constantes C independientes para cada clase en un SVM?
- a) Incorrecto. Las constantes son independientes por clase, por lo que no consideran los ejemplos duplicados de manera explícita.
 - b) **Correcto. Con constantes independientes es posible asignar importancias distintas a las dos clases.**
 - c) Incorrecto. El problema sigue manteniendo las mismas características de antes en cuanto a su esquema de resolución.
 - d) Incorrecto. Los multiplicadores de Lagrange son necesarios para manejar las restricciones y el usar dos constantes no tiene relación con esto.
6. ¿Cuáles de las siguientes aseveraciones sobre un kernel-SVM son correctas? (puedes seleccionar múltiples alternativas)
- a) Incorrecto. En el espacio transformado, un SVM se comporta como uno tradicional, desde el punto de su superficie de decisión.
 - b) **Correcto. Al hacer la proyección inversa, las rectas dejan cambiar de forma.**
 - c) **Correcto. Al usar superficies de decisión con más grados de libertad, existe mayor posibilidad de sobreajuste.**
 - d) Incorrecto. El kernel calcula un producto punto en el espacio transformado, no la transformación en sí.
7. *Entrenar un SVM sobre todo un conjunto de entrenamiento es lo mismo que entrenarlo sólo en los vectores de soporte de ese mismo conjunto.* Respecto a la afirmación anterior podemos afirmar que:
- a) **Correcto. Los vectores de soporte son aquellos que determinan el hiperplano separador y no el resto de los datos, por lo que podríamos entrenar solamente con ellos.**
 - b) Incorrecto. Recuerda que los vectores de soporte son aquellos que determinan el hiperplano separador y no el resto de los datos, por lo que podríamos entrenar solamente con ellos y obtener un mismo resultado.
8. Cuando trabajamos con SVM en muchas ocasiones podemos encontrarnos con problemas no linealmente separables. ¿Cómo podemos afrontar este problema para encontrar, de todos modos, un hiperplano separador? (Puedes seleccionar múltiples respuestas).
- a) **Correcto. Estas son conocidas como variables slack, utilizadas para las SVM soft margin, la incidencia de estas variables viene dada por el parámetro C de la optimización.**

- b) Incorrecto. El algoritmo de SVM trabaja con hiperplanos (rectas, planos, etc.), sin importar el espacio dimensional donde esto ocurra, por tanto, es imposible aumentar los grados de libertad de la superficie separadora. Se recomienda que en vez tratar de cambiar la superficie de clasificación, busquemos una solución en donde los datos se adapten a la superficie o bien el criterio para encontrarla sea menos exigente.
- c) Incorrecto. Esto solo generaría un nuevo problema entre la nueva clase y la antigua con la que solía haber un conflicto de no separabilidad, recuerda los conceptos de función de kernel y variables de slack vistas en la clase de esta semana.
- d) **Correcto. Este es el rol de la función de kernel.**

3. Redes Neuronales

9. El conjunto de datos HASYv2 contiene imágenes en blanco y negro de 369 íconos y símbolos matemáticos distintos, cada una con tamaño 30x30 píxeles.

Si quisiéramos entrenar una red neuronal para clasificar estas imágenes utilizando el valor de cada uno de sus píxeles como entrada, ¿cuántas neuronas necesitará tener la capa de entrada de la red?

- a) Incorrecto. Recuerda que el número de neuronas al comienzo depende del tamaño del vector de entrada.
- b) Incorrecto. Recuerda que el número de neuronas al comienzo no depende del número de clases, sino del tamaño del vector de entrada.
- c) **Correcto. Este corresponde al número total de píxeles en cada imagen, el número de neuronas al comienzo no depende del número de clases, sino del tamaño del vector de entrada.**
- d) Incorrecto. Recuerda que el número de neuronas al comienzo depende del tamaño del vector de entrada.

10. Respecto al número de neuronas de la red, podemos afirmar que posee:

- a) Incorrecto, recuerda que el número de neuronas de una capa se mide en paralelo (de forma vertical).
- b) **Correcto, notar que el número de neuronas se cuenta por capa (de forma vertical).**
- c) Incorrecto, recuerda que el número de neuronas de una capa se mide en paralelo (de forma vertical) y que la capa de salida corresponde a aquella que entrega la predicción de la red.
- d) Incorrecto, recuerda que la capa de salida corresponde a aquella que entrega la predicción de la red.

11. Con respecto a la entrada de la red, podemos afirmar que se debe tratar de:

- a) Incorrecto. Recuerda que neurona siempre recibirá un solo valor numérico de entrada, al tener 6 vectores de cualquier largo no podemos asegurar que la entrada contenga solamente 6 valores.
- b) Incorrecto. Recuerda que neurona siempre recibirá un solo valor numérico de entrada, una matriz de 6x6 posee 36 elementos distintos.
- c) **Correcto. Cada neurona siempre recibirá un solo valor numérico de entrada.**
- d) Incorrecto. Recuerda que neurona siempre recibirá un solo valor numérico de entrada, 6 imágenes de 28x28 contienen $6 \times 28 \times 28 = 4.704$ elementos distintos.

12. Con respecto a la salida de la red, podemos afirmar que se debe tratar de:

- a) **Correcto. La predicción corresponde a un vector con un valor por cada neurona en la capa de salida.**
 - b) Incorrecto. Recuerda que la predicción corresponde a un vector con un valor por cada neurona en la capa de salida, la red actual posee 3 neuronas en su capa de salida. Te recomendamos revisar la clase número sobre Fundamentos de Redes Neuronales para reforzar tu entendimiento la estructura de una red neuronal.
 - c) Incorrecto. Recuerda que la predicción corresponde a un vector con un valor por cada neurona en la capa de salida y no una matriz. Las dimensiones anteriores corresponden a las de la capa oculta de la red presentada.
 - d) Incorrecto. Recuerda que la predicción corresponde a un vector con un valor único por cada neurona en la capa de salida.
13. ¿Cómo propaga el algoritmo de backpropagation los pesos desde la capa de salida al resto de las capas?
- a) Incorrecto. Recuerda que el algoritmo de backpropagation, como su nombre lo indica, busca propagar la función de pérdida desde la capa de salida hacia la de entrada a modo de descender por su gradiente.
 - b) **Correcto. Esto a modo de ajustar cada peso en el sentido del gradiente de la función de pérdida.**
 - c) Incorrecto. Recuerda que backpropagation ajusta cada peso individualmente en base a la derivada parcial de la función de pérdida con respecto a él, esto garantiza de que cada peso pueda moverse en la dirección del gradiente.
 - d) Incorrecto. Recuerda que el algoritmo de backpropagation, como su nombre lo indica, busca propagar la función de pérdida desde la capa de salida hacia la de entrada a modo de descender por su gradiente.
14. ¿Con cuál(es) de los siguientes problemas puede trabajar un MLP de forma exitosa? (Puedes seleccionar múltiples alternativas)
- a) **Correcto. Dado que un MLP está compuesto por múltiples perceptrones, hace sentido que también pueda trabajar este tipo de problemas.**
 - b) **Correcto. Dado que un MLP está compuesto por múltiples perceptrones, hace sentido que también pueda trabajar este tipo de problemas.**
 - c) **Correcto. Esto es posible debido a que la función de activación utilizada en cada neurona es una de carácter no lineal.**
 - d) **Correcto. Esto es posible debido a que la función de activación utilizada en cada neurona es una de carácter no lineal.**

4. Aprendizaje Reforzado

15. ¿Cuál(es) de las siguientes opciones forma parte de un Proceso de Decisión de Markov? (Puedes seleccionar más de una alternativa)
- a) **Correcto. Un set de recompensas**
 - b) **Correcto. Un set de acciones**
 - c) **Correcto. Un set de probabilidades de transición entre estados**
 - d) Incorrecto. Un set de políticas forma parte del agente que modela el MDP

16. ¿Por qué utilizamos un factor de descuento en el cálculo del retorno G_t ?
- a) Incorrecto. Para penalizar acciones que sean distintas a la política actual.
 - b) **Correcto. Para ajustar la importancia que el agente les asignará a recompensas más lejanas en el tiempo.**
 - c) Incorrecto. Para evitar cambiar la política del agente constantemente.
 - d) Incorrecto. Para considerar en mayor o menor escala el próximo movimiento del oponente.
17. En el contexto de Q-Learning ¿En base a qué variables estima la función acción valor Q_π el valor del retorno? (Puedes seleccionar más de una alternativa)
- a) Incorrecto. A los estados anteriores s_0 a $s(t-1)$.
 - b) **Correcto. Al estado actual s_t .**
 - c) Incorrecto. Al valor esperado del estado anterior $G(t-1)$.
 - d) **Correcto. A la acción tomada a_t .**
18. ¿Cómo es computado el retorno esperado del agente en el siguiente estado?
- a) Incorrecto. Se estima la recompensa esperada de ejecutar la acción a_t mediante algún algoritmo de aprendizaje.
 - b) **Correcto. Se suman todas las recompensas esperadas en un horizonte de tiempo h , ponderando cada una por un factor de descuento.**
 - c) Incorrecto. Se aprende mediante prueba y error (en forma de la recompensa), almacenándose en la función Q_π .
 - d) Incorrecto. No se puede estimar el retorno esperado del agente.
19. ¿Cómo podemos definir a la política óptima de un agente?
- a) **Correcto. Aquella política que maximiza la recompensa a lo largo del tiempo.**
 - b) Incorrecto. Aquella política que minimiza la recompensa a lo largo del tiempo.
 - c) Incorrecto. Aquella política que minimiza el número de acciones para resolver un problema.
 - d) Incorrecto. Aquella política que minimiza la recompensa adversaria al jugar con un contrin-cante.