

## 1. Planteamiento del problema

1 / 1 punto

Este ejemplo está adaptado de una aplicación de producción real, pero con detalles disfrazados para proteger la confidencialidad.



Eres un investigador famoso en la Ciudad de Peacetopia. La gente de Peacetopia tiene una característica común: le tienen miedo a los pájaros. Para salvarlos, debes **construir un algoritmo que detecte cualquier pájaro que vuele sobre Peacetopia** y alerte a la población.

El Ayuntamiento te proporciona un conjunto de datos de 10.000.000 de imágenes del cielo sobre Peacetopia, tomadas de las cámaras de seguridad de la ciudad. Están etiquetados:

- $y = 0$ : No hay pájaro en la imagen
- $y = 1$ : hay un pájaro en la imagen

Tu objetivo es construir un algoritmo capaz de clasificar nuevas imágenes tomadas por cámaras de seguridad de Peacetopia.

Hay muchas decisiones que tomar:

- ¿Qué es la métrica de evaluación?
- ¿Cómo estructura sus datos en conjuntos de entrenamiento/desarrollo/prueba?

## Métrica de éxito

El Ayuntamiento te dice que quieren un algoritmo que

1. Tiene alta precisión
2. Se ejecuta rápidamente y toma poco tiempo para clasificar una nueva imagen.
3. Puede caber en una pequeña cantidad de memoria, por lo que puede ejecutarse en un pequeño procesador que la ciudad conectará a muchas cámaras de seguridad diferentes.

Nota : tener tres métricas de evaluación hace que le resulte más difícil elegir rápidamente entre dos algoritmos diferentes y reducirá la velocidad con la que su equipo puede iterar. ¿Verdadero Falso?

☒ Verdadero

☐ Falso

☒ **Correcto**

2. Después de más discusiones, la ciudad reduce sus criterios a:

1 / 1 punto

- "Necesitamos un algoritmo que nos permita saber que un pájaro está volando sobre Peacetopia con la mayor precisión posible".
- "Queremos que el modelo entrenado no tarde más de 10 segundos en clasificar una nueva imagen".
- "Queremos que el modelo quepa en 10 MB de memoria".

Si tuvieras los tres modelos siguientes, ¿cuál elegirías?

☐

Exactitud de la prueba	tiempo de ejecución	Tamaño de la memoria
97%	1 segundo	3MB

☐

Exactitud de la prueba	tiempo de ejecución	Tamaño de la memoria
99%	13 seg	9MB

☐

Exactitud de la prueba	tiempo de ejecución	Tamaño de la memoria
97%	3 seg	2MB

☒

Exactitud de la prueba	tiempo de ejecución	Tamaño de la memoria
98%	9 seg	9MB

☒ **Correcto**

¡Correcto! Tan pronto como el tiempo de ejecución sea inferior a 10 segundos, estará listo. Por lo tanto, simplemente puede maximizar la precisión de la prueba después de asegurarse de que el tiempo de ejecución sea <10 segundos.

3. Según las solicitudes de la ciudad, ¿cuál de las siguientes diría que es cierta?

1 / 1 punto

- ☒ La precisión es una métrica de optimización; el tiempo de ejecución y el tamaño de la memoria son métricas satisfactorias.
- ☐ La precisión es una métrica satisfactoria; el tiempo de ejecución y el tamaño de la memoria son una métrica de optimización.
- ☐ La precisión, el tiempo de ejecución y el tamaño de la memoria son métricas de optimización porque desea hacerlo bien en las tres.
- ☐ La precisión, el tiempo de ejecución y el tamaño de la memoria son métricas satisfactorias porque tiene que hacerlo lo suficientemente bien en las tres para que su sistema sea aceptable.

☒ **Correcto**

4. **Estructurando sus datos**

1 / 1 punto

Antes de implementar su algoritmo, debe dividir sus datos en conjuntos de entrenamiento/desarrollo/prueba. ¿Cuál de estos crees que es la mejor opción?

☐

Tren	desarrollador	Prueba
6,000,000	3,000,000	1,000,000

<input type="radio"/>	Tren	desarrollador	Prueba
	6,000,000	1,000,000	3,000,000
<input checked="" type="radio"/>	Tren	desarrollador	Prueba
	9,500,000	250,000	250,000
<input type="radio"/>	Tren	desarrollador	Prueba
	3,333,334	3,333,333	3,333,333

☒ **Correcto**  
Sí.

5. Después de configurar sus conjuntos de entrenamiento/desarrollo/prueba, el Ayuntamiento se encuentra con otras 1.000.000 de imágenes, denominadas "datos de los ciudadanos". Aparentemente, los ciudadanos de Peacetopia tienen tanto miedo a las aves que se ofrecieron como voluntarios para tomar fotografías del cielo y etiquetarlas, contribuyendo así con estas 1,000,000 de imágenes adicionales. Estas imágenes son diferentes de la distribución de imágenes que el Ayuntamiento te había dado originalmente, pero crees que podría ayudar a tu algoritmo.

1 / 1 punto

Tenga en cuenta que agregar estos datos adicionales al conjunto de entrenamiento hará que la distribución del conjunto de entrenamiento sea diferente de las distribuciones de los conjuntos de desarrollo y prueba.

¿La siguiente expresión es cierta o falsa?

"No debe agregar los datos de los ciudadanos al conjunto de entrenamiento, porque si la distribución del entrenamiento es diferente de los conjuntos de desarrollo y prueba, esto no permitirá que el modelo funcione bien en el conjunto de prueba".

- ☐ Verdadero  
☒ Falso

☒ **Correcto**

Falso es correcto: a veces necesitaremos entrenar el modelo con los datos disponibles, y su distribución puede no ser la misma que la de los datos que se producirán en producción. Además, agregar datos de entrenamiento que difieren del conjunto de desarrollo aún puede ayudar

al modelo a mejorar el rendimiento en el conjunto de desarrollo. Lo que importa es que el conjunto de desarrollo y prueba tengan la misma distribución.

6. Un miembro del Concejo Municipal sabe un poco sobre el aprendizaje automático y cree que debe agregar las imágenes de datos de 1,000,000 de ciudadanos al conjunto de prueba. Usted se opone porque:

1 / 1 punto

- ☐ Un conjunto de prueba más grande reducirá la velocidad de iteración debido al gasto computacional de evaluar modelos en el conjunto de prueba.
- ☐ Las imágenes de datos de 1.000.000 de ciudadanos no tienen un mapeo  $x \rightarrow y$  consistente como el resto de los datos (similar al ejemplo de precios de viviendas de la ciudad de Nueva York/Detroit de la conferencia).
- ☒ El equipo de prueba ya no refleja la distribución de datos (cámaras de seguridad) que más le interesan.

✓ Correcto

- ☒ Esto haría que las distribuciones de los conjuntos de desarrollo y prueba fueran diferentes. Esta es una mala idea porque no estás apuntando donde quieres golpear.

✓ Correcto

7. Entrena un sistema y sus errores son los siguientes (error = 100% de precisión):

1 / 1 punto

Error de conjunto de entrenamiento	4.0%
Error de conjunto de desarrollo	4.5%

Esto sugiere que una buena vía para mejorar el rendimiento es entrenar una red más grande para reducir el error de entrenamiento del 4,0 %. ¿Estás de acuerdo?

- ☐ Sí, porque tener un error de entrenamiento del 4,0 % muestra que tienes un alto sesgo.

- ☐ Sí, porque esto muestra que su sesgo es mayor que su varianza.
- ☐ No, porque esto muestra que su varianza es mayor que su sesgo.
- ☒ No, porque no hay información suficiente para saberlo.

☒ **Correcto**

8. Pide a algunas personas que etiqueten el conjunto de datos para averiguar qué es el rendimiento a nivel humano. Encontrará los siguientes niveles de precisión:

1 / 1 punto

Experto en observación de aves #1	0,3% de error
Experto en observación de aves #2	error del 0,5 %
Persona normal #1 (no un experto en observación de aves)	1,0% de error
Persona normal #2 (no un experto en observación de aves)	1,2% de error

Si su objetivo es que el "desempeño a nivel humano" sea un proxy (o estimación) del error de Bayes, ¿cómo definiría "desempeño a nivel humano"?

- ☐ 0.0% (porque es imposible hacerlo mejor que esto)
- ☒ 0.3% (precisión del experto #1)
- ☐ 0,4% (promedio de 0,3 y 0,5)
- ☐ 0,75% (promedio de los cuatro números anteriores)

☒ **Correcto**

9. ¿Con cuál de las siguientes afirmaciones está de acuerdo?

1 / 1 punto

- ☒ El rendimiento de un algoritmo de aprendizaje puede ser mejor que el rendimiento a nivel humano, pero nunca puede ser mejor que el error de Bayes.

- ☐ El rendimiento de un algoritmo de aprendizaje nunca puede ser mejor que el rendimiento a nivel humano, pero puede ser mejor que el error de Bayes.
- ☐ El rendimiento de un algoritmo de aprendizaje nunca puede ser mejor que el rendimiento a nivel humano ni mejor que el error de Bayes.
- ☐ El rendimiento de un algoritmo de aprendizaje puede ser mejor que el rendimiento a nivel humano y mejor que el error de Bayes.
- ☒ **Correcto**

10. Encuentra que un equipo de ornitólogos que debaten y discuten una imagen obtiene un rendimiento aún mejor del 0,1%, por lo que lo define como "rendimiento a nivel humano". Después de trabajar más en su algoritmo, termina con lo siguiente:

1 / 1 punto

Desempeño a nivel humano	0.1%
Error de conjunto de entrenamiento	2.0%
Error de conjunto de desarrollo	2.1%

Según la evidencia que tiene, ¿cuáles dos de las siguientes cuatro opciones parecen ser las más prometedoras para probar? (Marque dos opciones).

- ☒ Entrena a un modelo más grande para intentar hacerlo mejor en el set de entrenamiento.

☒ **Correcto**

- ☐ Obtenga un conjunto de entrenamiento más grande para reducir la variación.
- ☐ Intente aumentar la regularización.
- ☒ Intente disminuir la regularización.

☒ **Correcto**

11. También evalúa su modelo en el conjunto de prueba y encuentra lo siguiente:

1 / 1 punto

Desempeño a nivel humano	0.1%
Error de conjunto de entrenamiento	2.0%
Error de conjunto de desarrollo	2.1%
Error de conjunto de prueba	7.0%

¿Qué significa esto? (Marque las dos mejores opciones).

- ☐ Tienes un ajuste inadecuado para el conjunto de desarrollo.
- ☒ Deberías intentar obtener un conjunto de desarrollo más grande.

✓ **Correcto**

- ☒ Ha sobreajustado al conjunto de desarrollo.

✓ **Correcto**

- ☐ Debería obtener un conjunto de prueba más grande.

12. Después de trabajar en este proyecto durante un año, finalmente logras:

1 / 1 punto

Desempeño a nivel humano	0.10%
Error de conjunto de entrenamiento	0.05%
Error de conjunto de desarrollo	0.05%

¿Qué puedes concluir? (Marque todo lo que corresponda.)

- ☐ Esta es una anomalía estadística (o debe ser el resultado de ruido estadístico) ya que no debería ser posible superar el rendimiento a nivel humano.
- ☒ Si el conjunto de prueba es lo suficientemente grande para que la estimación del error del 0,05% sea precisa, esto implica que el error de Bayes es  $\leq 0.05$

✓ **Correcto**

- ☐ Con solo un 0,09 % más de progreso por hacer, debería poder cerrar rápidamente la brecha restante al 0 %.



- ☒ Ahora es más difícil medir el sesgo evitable, por lo que el progreso será más lento en el futuro.

☒ Correcto

**13.** Resulta que Peacetopia ha contratado a uno de sus competidores para construir un sistema también. Tanto su sistema como el de la competencia ofrecen sistemas con aproximadamente el mismo tiempo de ejecución y el mismo tamaño de memoria. ¡Sin embargo, su sistema tiene mayor precisión! Sin embargo, cuando Peacetopia prueba sus sistemas y los de su competidor, concluyen que en realidad les gusta más el sistema de su competidor, porque a pesar de que tiene una precisión general más alta, tiene más falsos negativos (no puede dar la alarma cuando un pájaro está en el aire) . ¿Qué debes hacer?

1 / 1 punto

- ☐ Mire todos los modelos que ha desarrollado durante el proceso de desarrollo y encuentre el que tenga la tasa de error de falsos negativos más baja.
- ☐ Pida a su equipo que tenga en cuenta tanto la precisión como la tasa de falsos negativos durante el desarrollo.
- ☒ Reconsidere la métrica adecuada para esta tarea y pídale a su equipo que se adapte a la nueva métrica.
- ☐ Elija la tasa de falsos negativos como la nueva métrica y utilícela para impulsar todo el desarrollo futuro.

☒ Correcto

**14.** ¡Ha vencido fácilmente a su competidor, y su sistema ahora está implementado en Peacetopia y está protegiendo a los ciudadanos de las aves! Pero en los últimos meses, una nueva especie de ave ha estado migrando lentamente al área, por lo que el rendimiento de su sistema se degrada lentamente porque sus datos se prueban en un nuevo tipo de datos.

1 / 1 punto



Tienes solo 1.000 imágenes de la nueva especie de ave. La ciudad espera un mejor sistema de su parte dentro de los próximos 3 meses. ¿Cuál de estos debes hacer primero?

- ☒ Use los datos que tiene para definir una nueva métrica de evaluación (usando un nuevo conjunto de desarrollo/prueba) teniendo en cuenta las nuevas especies, y utilícelas para impulsar el progreso de su equipo.
- ☐ Coloque las 1000 imágenes en el conjunto de entrenamiento para tratar de hacerlo mejor con estas aves.
- ☐ Pruebe el aumento de datos/síntesis de datos para obtener más imágenes del nuevo tipo de ave.
- ☐ Agregue las 1000 imágenes a su conjunto de datos y reorganícelas en una nueva división de entrenamiento/desarrollo/prueba.

☒ **Correcto**

**15.** El Ayuntamiento cree que tener más Gatos en la ciudad ayudaría a ahuyentar a las aves. Están tan contentos con su trabajo en el detector de pájaros que también lo contrataron para construir un detector de gatos. (Wow, los detectores de gatos son increíblemente útiles, ¿no?) Debido a años de trabajo en detectores de gatos, tiene un conjunto de datos tan grande de 100,000,000 **0.75 / 1 punto**

de imágenes de gatos que el entrenamiento con estos datos toma alrededor de dos semanas. ¿Con cuál de las afirmaciones estás de acuerdo? (Marque todas las que estén de acuerdo).

- ☒ Si 100 000 000 de ejemplos son suficientes para construir un detector de gatos lo suficientemente bueno, podría ser mejor entrenar con solo 10 000 000 de ejemplos para obtener una  $\approx 10$  veces más rápido en la ejecución de experimentos, incluso si cada modelo funciona un poco peor porque está entrenado con menos datos.

☒ **Correcto**

- ☒ Necesitar dos semanas para entrenar limitará la velocidad a la que puede iterar.

☒ **Correcto**

- ☐ Habiendo construido un buen detector Bird, debería poder tomar el mismo modelo e hiperparámetros y simplemente aplicarlo al conjunto de datos Cat, por lo que no hay necesidad de iterar.
- ☐ Comprar computadoras más rápidas podría acelerar la velocidad de iteración de sus equipos y, por lo tanto, la productividad de su equipo.

No seleccionaste todas las respuestas correctas