

1)

- que es backpropagation?

El Backpropagation (o propagación hacia atrás) es el algoritmo fundamental utilizado para el entrenamiento de redes neuronales multicapa. Su objetivo principal es encontrar el mínimo de la función de error para ajustar los pesos del modelo de manera óptima.

- que es softmax?

Función exponencial normalizada. Se utiliza como función de activación de la capa de salida en modelos de clasificación, interpretándose como scoring, según el modelo, de pertenecer a dicha clase.

- que son los optimizadores? Menciona 3

Un optimizador es el algoritmo que actualiza los parámetros del modelo a partir de los gradientes de la función de pérdida (determina la regla de actualización y su dinámica). Es una implementación concreta del algoritmo de backpropagation

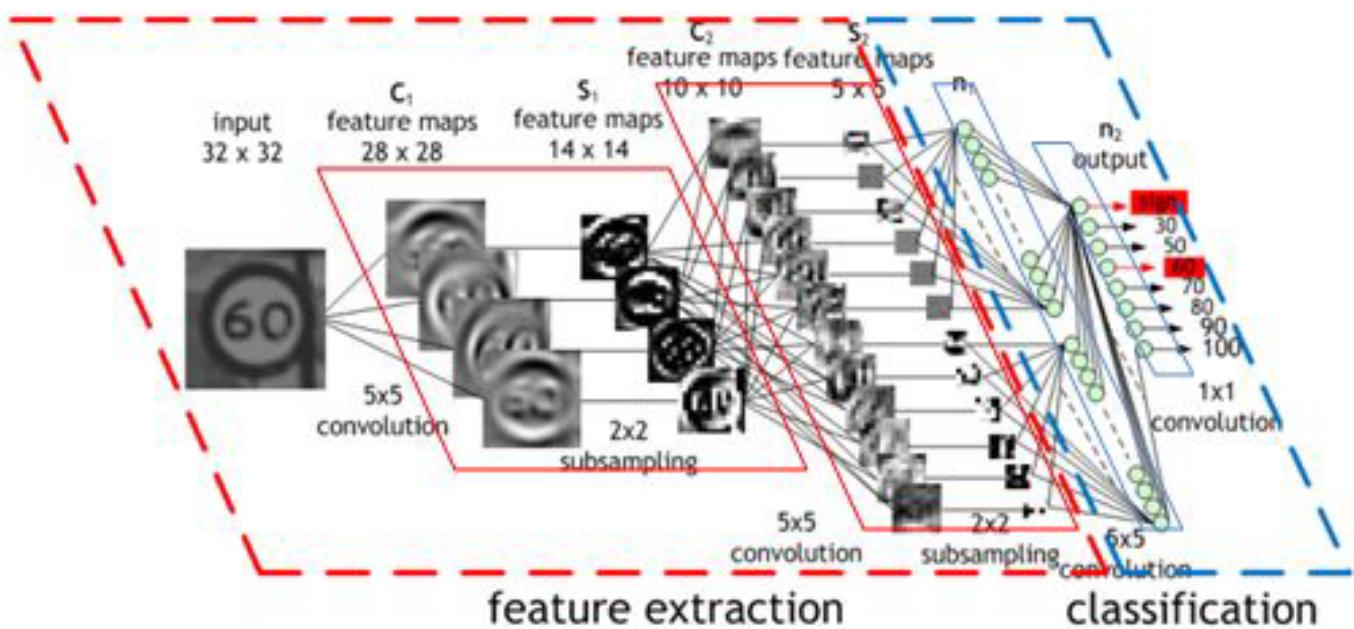
- Momentum
- Nesterov
- RMSprop
- AdaGrad
- Adam
- Nadam

- que es RAG?

Es el proceso de optimización de la salida de un modelo lingüístico de gran tamaño, de modo que haga referencia a una base de conocimientos autorizada fuera de los orígenes de datos de entrenamiento antes de generar una respuesta.

2)

- diagramar una red convolucional



- para que usaría una red recurrente?

Las redes neuronales recurrentes (RNN) se utilizan principalmente para procesar datos secuenciales donde el orden de la información es fundamental para entender su significado. A diferencia de otras redes, estas poseen un bucle de retroalimentación (feedback-loop) que les permite mantener una especie de memoria, incorporando conocimientos de entradas anteriores para influir en el procesamiento actual.

- cómo hace bayes para calcular la probabilidad de que una palabra esté en un documento?



Naïve Bayes - clasificación de texto

¿Cuál es la probabilidad de que el documento d pertenezca a la clase c ?

Parafraseado como probabilidad condicional:

Dado el documento d , ¿cuál es la probabilidad de que pertenezca a c ?
 $= P(c | d)$

Y por el teorema de Bayes se puede plantear lo siguiente:

$$P(c | d) = \frac{P(d | c) P(c)}{P(d)}$$



Naïve Bayes - clasificación de texto

$C_{\text{map}} = \underset{\mathbf{C}}{\operatorname{argmax}} P(c | d)$ para $c \in \mathbf{C}$, (el conjunto de todas las clases).

Por Bayes:

En el denominador nos queda $P(d)$ como una constante. Lo eliminamos.

$$C_{\text{map}} = \underset{\mathbf{C}}{\operatorname{argmax}} \frac{P(d | c) P(c)}{P(d)}, \quad c \in \mathbf{C}$$

$$C_{\text{map}} = \underset{\mathbf{C}}{\operatorname{argmax}} P(d | c) P(c)$$

- 3 problemas que solucionan las redes recurrentes dependiendo de la entrada y la salida

- Entrada: **secuencia** / Salida: **secuencia** : Clasificar videos cuadro a cuadro, Series temporales.
- Entrada: **secuencia** / Salida: **singular** : Clasificación de documentos. Análisis de Sentimientos.
- Entrada: **singular** / Salida: **secuencia** : Etiquetamiento: Utilizado para el etiquetamiento de imágenes, donde a partir de una imagen fija la red genera una secuencia de palabras que funciona como título o descripción

3)

- que son los aspectos?

En el análisis de sentimientos, los aspectos (también llamados atributos u objetos de sentimiento) representan el componente específico de un producto o servicio que está siendo evaluado.

- Función: Permiten detectar más de un sentimiento en una misma frase. Por ejemplo: "La comida era excelente pero el servicio pésimo". Aquí, "comida" y "servicio" son los aspectos con polaridades opuestas.

Ejemplos: Rating u Opiniones

- que son los embeddings?

Los embeddings son una técnica que consiste en mapear palabras a vectores reales. Es la forma en que los modelos de ciencia de datos pueden procesar el texto, ya que no trabajan con caracteres sino con optimizaciones numéricas.

MAS LARGO:

- Como su nombre lo indica, los vectores de palabras son vectores que se utilizan para representar palabras y también se pueden considerar como vectores de características o representaciones de palabras.
- La técnica de mapear palabras a vectores reales se llama Embedding de palabras. En los últimos años, los embedding de palabras se han convertido gradualmente en el conocimiento básico del procesamiento del lenguaje natural.

- que es el lexicon de emociones?

Un lexicón es esencialmente un diccionario o lista de palabras que tienen asignada una categoría o carga emocional. Se utilizan para determinar la polaridad de un texto comparando sus palabras con estas listas predefinidas. Existen varios tipos conocidos:

- LIWC (Linguistic Inquiry and Word Count): Contiene unas 2.300 palabras, posee clasificaciones complejas y tiene la ventaja de soportar el idioma español.

- The General Inquirer: Uno de los más antiguos, con categorías de palabras "positivas" y "negativas".
- SentiWordNet: Basado en WordNet, donde una misma palabra puede tener valores positivos, negativos y objetivos simultáneamente.
- MPQA Subjectivity Cues Lexicon: Además de la polaridad, indica la intensidad de la palabra (si es "fuerte" o "débil").

- que son las redes SOM?

SOMs son redes no supervisadas para visualización y clustering de datos de alta dimensión en un mapa 2D. Útiles para explorar estructuras, agrupar datos, y crear mapas que muestran similitudes entre observaciones (por ejemplo, segmentación de clientes, análisis exploratorio).

4)

- diferencias entre xgboost y RandomForest

La diferencia principal entre XGBoost y Random Forest radica en la técnica de ensamble que utilizan: Random Forest se basa en Bagging (entrenamiento paralelo e independiente), mientras que XGBoost utiliza Boosting (entrenamiento secuencial donde cada paso corrige al anterior)

1. Metodología de Ensamble

- Random Forest (Bagging): Construye múltiples árboles de decisión de forma independiente utilizando subconjuntos aleatorios de datos (bootstrap) y promedia sus resultados mediante votación.
- XGBoost (Boosting): Crea una cadena de árboles de forma secuencial. Cada nuevo árbol se entrena para corregir los errores (residuos) cometidos por los árboles previos.

2. Entrenamiento y Performance

- Paralelismo: En Random Forest, los árboles pueden entrenarse en paralelo porque no dependen entre sí. En XGBoost, la creación de un modelo depende del anterior, aunque su implementación está optimizada para el uso eficiente del hardware (GPU) y Big Data.
- Complejidad de los modelos: Random Forest suele utilizar árboles completos y profundos. XGBoost utiliza árboles de profundidad fija (a menudo

limitados) que actúan como estimadores débiles que se fortalecen en conjunto.

3. Manejo de Sesgo y Varianza

- Objetivo de Random Forest: Se enfoca principalmente en reducir la varianza y evitar el overfitting promediando modelos de bajo sesgo y alta varianza.
- Objetivo de XGBoost: Busca un equilibrio optimizando tanto el sesgo (corrigiendo errores de suposiciones erróneas) como la varianza (mediante regularización).

4. Regularización y Control de Errores

- Regularización: XGBoost incluye parámetros de regularización avanzada como λ (lambda) y γ (gamma) para penalizar la complejidad del modelo y evitar el sobreajuste, algo que el Random Forest tradicional no maneja de la misma manera.
- Poda: XGBoost realiza una poda de los árboles basada en el parámetro gamma, eliminando nodos donde la ganancia de información no supera un umbral específico.

Cuadro Comparativo de Referencia

- Características principales del boosting

El Boosting es una técnica de ensamble que consiste en entrenar predictores de manera secuencial, donde cada nuevo modelo intenta corregir los errores cometidos por los modelos anteriores. A diferencia del Bagging, que entrena modelos de forma independiente, el Boosting busca convertir una serie de estimadores débiles (weak learners) en un clasificador fuerte.

- 2 métodos de xgboost para evitar overfitting

XGBoost utiliza principalmente la regularización λ (lambda) y la poda mediante el parámetro γ (gamma) para controlar la complejidad del modelo y evitar el sobreajuste o overfitting.

- que es el método cascada?

El método de Cascading (o cascada) es un tipo de ensamble híbrido que utiliza un enfoque secuencial o de "multi-estado" para realizar predicciones. A diferencia de otras técnicas como Voting o Stacking, los datos pasan sucesivamente de un modelo a otro, donde cada "capa" está compuesta por un único modelo.

Ejemplo de Aplicación: Detección de Fraude

Ventajas y Desventajas

- Precisión: Generalmente logran predicciones más precisas que los modelos individuales y ofrecen un buen compromiso entre sesgo y varianza.
- Overfitting: Un riesgo crítico es que cascadas muy profundas pueden producir sobreajuste o overfitting.
- Complejidad: Se pierde la interpretabilidad del modelo (se vuelve una "caja negra") y requiere un costo computacional más elevado.

5)

- de que sirve la reducción de dimensiones?

La reducción de dimensionalidad sirve principalmente para proyectar datos que se encuentran en un espacio de alta complejidad hacia una dimensión menor (generalmente 2D o 3D), intentando preservar características esenciales como las distancias y las correlaciones entre los puntos.

- Visualización de datos para entender su distribución
 - detección de patrones inherentes a simple vista
- Reducción del ruido
- Aceleración de los tiempos de entrenamiento de un modelo
- Compresión de la información
- Presentación de resultados a interesados (quienes no siempre conocen de ciencia de datos)

- para que usaría un scree plot con apl?

Un Scree Plot es una herramienta visual que se utiliza en el Análisis de Componentes Principales (PCA) para representar el porcentaje de variación que aporta cada Componente Principal (PC) al conjunto de datos total.

- que método usar si sospechamos que los datos están en una variedad del espacio

ISOMAP

- que una definición de t-sne y te preguntan a qué método pertenece la definición

PCA:

- El análisis de componentes principales, o PCA, nos sirve para identificar si hay agrupamiento de datos en el espacio de entrada.
- Podemos identificar correlaciones, clusters o bien entender cuán dispersos están los datos y sobre todo, sobre qué ejes o variables.
- Es útil especialmente cuando no podemos representar el espacio de entrada sobre un eje cartesiano.

T-SNE: Proyectar a un espacio menor para preservar los clusters.

MDS: Preservar las distancias entre puntos

ISOMAP: Variedad del espacio

6)

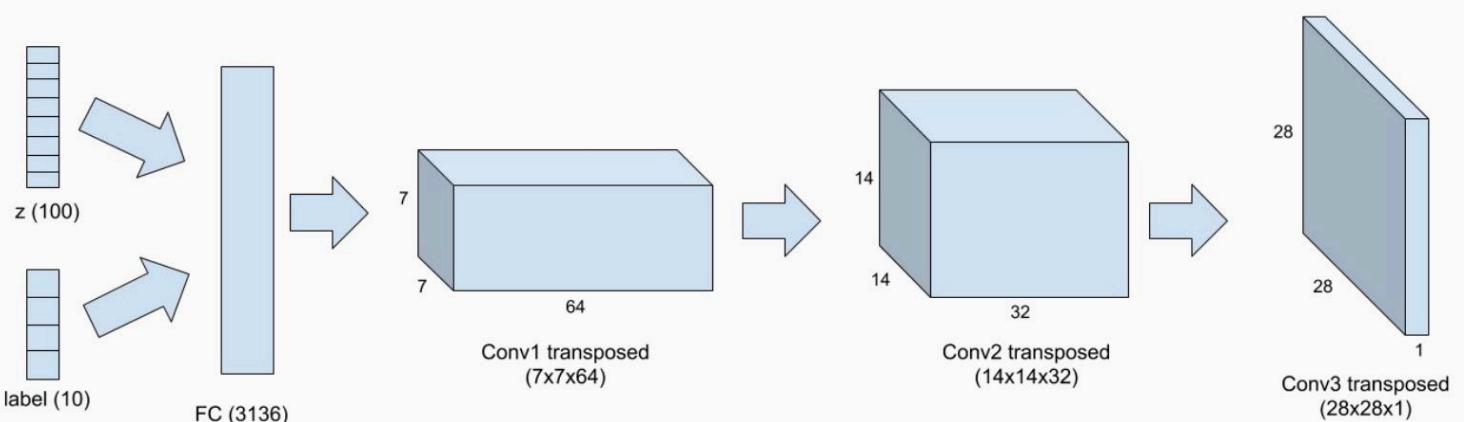
- que es el margen en una svm

En el contexto de una Support Vector Machine (SVM), el margen se define como la distancia mínima que existe entre cualquier observación del conjunto de datos y el umbral de decisión o límite.

- Sacado del pdf: La distancia mínima entre cualquier observación y el umbral se llama: margen

Un clasificador de Margen Máximo es muy sensible a los datos de entrenamiento (bajo sesgo), pero puede fallar con datos nuevos (alta varianza).

- diagramar una red neuronal para clasificar imágenes de 28x28 según que dígito es la imagen



- si se quiere hacer un modelo para predecir fraudes, que métrica usar

(MEJOR RESPUESTA): Cascading (Cascada): Debido a que se necesita una alta certeza para definir que algo no es fraude, se suelen usar ensambles de tipo cascada. En este enfoque, si un modelo inicial no tiene una probabilidad de "no fraude" superior al 0.99 (99%), la transacción pasa a un modelo más complejo o a una revisión manual.

(RESPUESTA DUDOSA): Para predecir fraudes, la métrica más importante suele ser el Recall (Exhaustividad), aunque la elección final depende del objetivo de negocio y del costo asociado a los errores.

En un problema de detección de fraudes, las métricas se definen a partir de la Matriz de Confusión, que organiza los resultados en cuatro categorías: Verdaderos Positivos (TP), Verdaderos Negativos (TN), Falsos Negativos (FN) y Falsos Positivos (FP).

1.

Te dan un cuadro con cinco documentos que pertenecen a una clase (POS, NEG y uno no se sabe) y te piden calcular el vocabulario, probabilidades de que un documento sea POS y de NEG y las probabilidades del documento que no tiene clase de de POS o NEG

2.

a. Te preguntan que son los aspectos en el análisis de sentimientos y que des ejemplos

En el análisis de sentimientos, los aspectos (también llamados atributos u objetos de sentimiento) representan el componente específico de un producto o servicio que está siendo evaluado.

• Función: Permiten detectar más de un sentimiento en una misma frase. Por ejemplo: "La comida era excelente pero el servicio pésimo". Aquí, "comida" y "servicio" son los aspectos con polaridades opuestas.

Ejemplos: Rating u Opiniones

b. Te dan una oración y te piden que hagas un Bag of Words

c. Te preguntan que son las StopWords y te piden ejemplos.

Las StopWords (o palabras de parada) son términos que se filtran, omiten o eliminan durante el procesamiento de un texto por considerarse poco informativos para ciertas tareas de aprendizaje automático.

Según los ejemplos gráficos de los documentos, estas son algunas de las palabras que se suelen filtrar:

- Artículos: la, el, los, un, una.
- Preposiciones: por, en, sin, con.
- Conjunciones: y, que.
- Pronombres: nos.

3.

a. Te preguntan para que sirve la reducción de dimensionalidad

La reducción de dimensionalidad sirve principalmente para proyectar datos que se encuentran en un espacio de alta complejidad hacia una dimensión menor (generalmente 2D o 3D), intentando preservar características esenciales como las distancias y las correlaciones entre los puntos.

- Visualización de datos para entender su distribución
 - detección de patrones inherentes a simple vista
- Reducción del ruido
- Aceleración de los tiempos de entrenamiento de un modelo
- Compresión de la información
- Presentación de resultados a interesados (quienes no siempre conocen de ciencia de datos)

b. Te preguntaban que técnica usarías si sabes que la dimensión está en una variedad del espacio

ISOMAP

c. Te preguntaban que técnica usarías si deseas preservar los clusters

T-SNE

d. Te preguntaban que técnicas usarías para preservar las distancias.

MDS

4.

a. Te preguntan las diferencias entre Stacking y Voting

Característica	Voting	Stacking
Enfoque	Multi-experto (simultáneo)	Multi-experto (con meta-modelo)
Decisión final	Votación por mayoría o ponderada	Aprendida por un meta-modelo
Niveles	Un solo nivel de modelos base	Puede tener múltiples capas apiladas
Concepto base	Sabiduría de las multitudes	Meta-aprendizaje

b. Te preguntan las diferencias entre Bagging y Boosting

La diferencia fundamental entre Bagging y Boosting radica en la forma en que se entrena los modelos y el objetivo principal que persiguen dentro de un esquema de ensambles. Mientras que Bagging busca reducir la varianza entrenando modelos en paralelo, Boosting busca reducir el error entrenando modelos de forma secuencial.

Característica	Bagging	Boosting
Orden de modelos	Independiente / Paralelo	Secuencial / Dependiente
Muestra de datos	Bootstrap (muestreo con reemplazo)	Ponderación de errores previos
Objetivo principal	Reducir Varianza	Reducir Sesgo y Error
Votación	Simple (mayoría)	Ponderada (por precisión)
Ejemplo clásico	Random Forest	AdaBoost, XGBoost

c. Te piden dos mecanismos que usa XGBoost para reducir el sobreajuste

XGBoost utiliza principalmente la regularización λ (lambda) y la poda mediante el parámetro γ (gamma) para controlar la complejidad del modelo y evitar el sobreajuste o overfitting.

5.

a. Te preguntan qué son las redes GAN y su uso (también hacer su diagrama)

Las Redes Generativas Adversarias (conocidas como GAN, por sus siglas en inglés) son un tipo de arquitectura de aprendizaje profundo no supervisado diseñada para generar datos nuevos que sean similares a los datos de entrenamiento.

Características:

-Son redes NO-Supervisadas

- No arman clusters
- No clasifican datos
- No resuelven problemas de regresión

b. Te preguntan qué son las celdas GRU y LTSM

Las celdas LSTM y GRU son variantes avanzadas de las neuronas recurrentes tradicionales, diseñadas específicamente para solucionar el problema de la memoria a corto plazo y el desvanecimiento del gradiente en las redes neuronales recurrentes (RNN).

A diferencia de las celdas simples, estas arquitecturas permiten que la red aprenda a recordar información relevante durante períodos prolongados y decida qué datos descartar.

Característica	LSTM	GRU
Complejidad	Mayor (3 puertas).	Menor (versión simplificada).
Estados	Separados (corto y largo plazo).	Fusionados en un solo vector.
Uso de recursos	Más lenta de entrenar.	Más rápida y eficiente.

c. Te preguntan qué son los embeddings

Los embeddings son una técnica que consiste en mapear palabras a vectores de números reales. Es la forma en que los modelos de ciencia de datos pueden procesar el texto, ya que no trabajan con caracteres sino con optimizaciones numéricas.

d. Te preguntan qué son los autoencoders

Restricted Boltzmann Machines (RBM) y Deep Belief Nets pueden usarse como Autoencoders

- Un autoencoder aprende a producir a la salida exactamente la misma información que recibe a la entrada

- Entrada y salida igual número de neuronas
- Compresor
- Reducción de dimensionalidad
 - 28x28 pixeles tiene 750 => solo 30 números
- Eliminación de ruido

6.

a. Te preguntan que hace el SVM cuando se enfrenta a un caso que no es linealmente separable

Cuando una Máquina de Vectores de Soporte (SVM) se enfrenta a un conjunto de datos que no es linealmente separable —es decir, que no puede dividirse con una simple línea o plano—, utiliza una técnica para llevar los datos a una dimensión mayor donde sí sea posible encontrar un separador.

b. Te preguntan si un Perceptrón simple puede resolver problemas que no son linealmente separables

Solo puede separar clases linealmente separables. Si las clases no son separables por una hiperplano (p. ej. XOR), el perceptrón no converge.

c. Te preguntan que método matemático usa las redes neuronales para reducir el error

El método matemático fundamental que utilizan las redes neuronales para reducir el error es el Backpropagation (o propagación hacia atrás), el cual se apoya principalmente en la técnica del descenso por gradiente.

Utiliza la regla de la cadena para calcular gradientes de la función de pérdida respecto a los pesos en una red multicapa y luego aplica descenso del gradiente (o variantes) para actualizar los pesos.