MIT-ML: Módulo 7

1 Apps TDD

TDD: seleccionar la mejor accion entre varias posibles en un entorno de incertidumbre

Decisión Optima: Dependen del Objetivo Final del proyecto, de mediano y largo plazo

Preguntas clave para resolver un problema

- ¿Con qué datos contamos?
- ¿En qué cuadrante de la clasificación de contextos de tomas de decisiones encaja?
- ¿Cuál es el objetivo?
- ¿De qué acciones disponemos?

4 Modelos

1) Modelo Predictivo Control (MPC)

Dinámica Entorno	Fija	
Disponib Info	Alta	Baja

Cuando el entorno apenas varía y se dispone de muchos datos, se recurre al modelo predictivo de control. Sería el caso de:

- Fijación de precios.
- Selección del portfolio.

2) Multi-Armed Bandit (MAB)

Dinámica Entorno	Fija	
Disponib Información	Alta	Baja

Cuando el entorno es relativamente estable y se cuenta de grandes volúmenes de datos, es utiles el multi-armed bandit. Correspondería a situaciones como:

- Gestión del inventario.
- Operaciones automatizadas financiero.
- Selección de rutas en una ciudad nueva.

3) Proceso Decision Markov (MDP)

Dinámica Entorno	Alta	Baja
Disponib Información	Alta	Baja

Cuando el entorno es muy dinámico y hay muchos datos disponibles, la solución óptima es el marco de decisiones de Markov. Ejemplificarlas sería casuística:

- Sugerencias en tiempo real.
- Selección de ruta en una ciudad concurida.

4) Aprendizaje x Refuerzo (AxR)

Dinámica Entorno	Alta	Baja
Disponib Información	Alta	Baja

Cuando el entorno fluctúa mucho y no se dispone de datos historicos, nos servirnos del aprendizaje por refuerzo. Se aplicaría a casos como:

- Gestion de posibles.
- Operaciones automatizadas.
- Selección de ruta con obstáculos.

Seleccionar a asesor que dé consejos más rentables

Punto de Partida: Cada día debemos:
- Elegir una sola acción en la que invertir
- Buscar asesoramiento financiero entre diversas personas.
- Decidir comprar, vender o mantenerme con base en estos consejos.

Objetivo: Nos proponemos localizar el mejor asesor y seguir sus consejos para maximizar nuestras ganancias.

Desafío: No disponemos de ningún dato a priori sobre quién puede asesorarnos mejor.

Imagina que cada día:
1. Elegirás solo una acción en la que invertir
2. Buscas asesoramiento financiero de expertos financieros o amigos
3. Decides si comprar, vender o no hacer nada basándote en sus consejos

Objetivo: Seguir los consejos del mejor asesor
Desafío: No sabes quién es el mejor

Explorar para que la calidad de las sugerencias de cada asesor hayan sido probadas y explotar los buenos asesores.

El **multi-armed bandit** se puede utilizar para seleccionar al mejor asesor financiero

Dinámica de entorno	Fija	
Disponibilidad de información	Alta	Baja

- El entorno se mantiene fijo
- Se recopila información cada vez que se recomienda asesor ya que no hay información previa

Este sistema permite tomar decisiones óptimas desde el primer momento.
Selecciona la respuesta correcta y haz clic en "Enviar".
- Verdadero
- Falso

En fase explotación....

Usar el **multi-armed bandit** para la correcta **exploración y explotación** de **compensaciones** puede ayudar a determinar la **solución** óptima, lo que equivale a tener **conocimiento previo** sobre el **mejor asesor**

Caso 2: Multi-Armed Bandit

2) Multi-Armed Bandit (MAB)

Dinámica Entorno	Fija	
Disponib Info	Alta	Baja

3.1 Selec Asesor

Nosotros tomamos las decisiones de inversión como cartera

En este caso, vamos a estructurar nuestra cartera de inversión. Compraremos que activos son más rentables e invertiremos en ellos durante cierto tiempo. No será hasta pasado un tiempo prudencial que evaluaremos el rendimiento de nuestra cartera de inversiones.

Suponiendo que:
- Tenes un modelo predictivo
- Mediante el modelo predictivo, tomas una decisión de cartera en tu periodo de decisión
- Durante el periodo de decisión, no hay gestión activa
- Revisas la decisión para el siguiente periodo de decisión

- **Tienes un modelo predictivo que predice los resultados de tu portafolio de inversión durante un periodo de tiempo con una desviación y distribución estándar**
- **El entorno no cambia durante el periodo de decisión**

Supuestos

1) Modelo Predictivo Control (MPC)

Dinámica Entorno	Fija	
Disponib Info	Alta	Baja

3.2 Selec Portfolio

Define el objetivo de control del modelo predictivo del control

- Determinar las limitaciones
- Determinar el objetivo

Paso 1: Predicción
- Discriminar el modelo predictivo de control mediante el portafolio durante un periodo de tiempo

Paso 2: Optimización

Ordena los pasos que deberías seguir para realizar la selección del portfolio.
Selecciona y ordena las respuestas para ordenarlas correctamente, después haz clic en "Enviar".
- Comprender el modelo predictivo.
- Determinación de objetivos.
- Optimización del modelo predictivo.

Usa un error al Modelo predictivo de control mienta que te permite mejorar mienta y repite la decisión

Dinámica de entorno	Fija	
Disponibilidad de información	Alta	Baja

- Disponemos de un modelo predictivo fuerte
- El entorno se mantiene fijo durante el periodo de decisión
- Se toman decisiones solo una vez

Caso 1: Modelo Predictivo de Control

Nuestro objetivo es automatizar las operaciones financieras, para que nuestro sistema de toma de decisiones ajuste la estructura de la cartera a la situación del mercado.

OPERACIONES AUTOMATIZADAS

Sistema que toma las decisiones de inversión

¿Con qué limitaciones podría contar este proceso?
Selecciona la respuesta correcta y haz clic en "Enviar".
- Precisa tiempos de procesamiento considerables.
- Precisa de software muy complejo.
- Precisa conocer muchos datos de entrenamiento.

Usar la información que proporcione la política por aprendizaje por refuerzo
Incrementar nuestra información y la identificación del modelo
Optimizar las decisiones en tiempo real
Creación de una política final

Aprendizaje por refuerzo para operaciones automatizadas

Dinámica de entorno	Alta	Baja
Disponibilidad de información	Alta	Baja

- El entorno cambia constantemente
- Nivel de información alto

Para las **operaciones automatizadas**, es necesario **maximizar el objetivo** a largo plazo mientras que las **microdecisiones** se toman en **tiempo real**

3.3 Ops. Automatizadas

2 Apps Retail

1) Modelo Predictivo Control (MPC)

Dinámica Entorno	Fija	
Disponib Info	Alta	Baja

2.1 Fijación Precios

¿Cómo fijas el precio de los productos?

Una vez producido un producto se debe establecer el precio al cual ofertarlo

Precio bajo → Pérdida de ingresos
Precio alto → Pérdida de clientes

Dilema en fijacion de precios: equilibrar 2 necesidades

- Precio debe ser competitivo, lo suficientemente bajo para que sea atractivo a los clientes potenciales
- Precio debe ser rentable, generar suficientes beneficios

Caso 1: Modelo Predictivo de Control MPC

Orientado a la mejora de decisiones futuras. Aprende el entorno a partir de datos históricos

Asumimos que nuestra empresa ha generado y recopilado una gran cantidad de información de benchmark sobre ventas y estado del negocio y que esta situación es relativamente estática.

Cuál es la situación?

Cuál es el punto más adecuado para maximizar ingresos, beneficios o ambos ?

Modelo Predictivo de Control para fijación de precios

Paso 1: Predicción

Objetivo: Estimar la demanda del producto a un precio a través de todos los canales de venta

La red neuronal es la mejor opción para predecir la fijación de precios en la venta al por menor porque:
- La venta al por menor **posee** un gran volumen de datos
- Son datos **no estructurados**

Paso 2: Optimización

Existen softwares de optimización y soluciones que nos pueden ayudar a resolver el problema

El modelo predictivo de control puede ayudar a **determinar el precio correcto** para lograr tu objetivo

¿Qué predicción se utiliza como input para optimizar un determinado objetivo?
Selecciona la respuesta correcta y haz clic en "Enviar".
- La maximización de beneficios.
- La probabilidad incremental de compra.
- La predicción de beneficios en caso de que un producto se venda o de pérdidas si no se vende.

2) Multi-Armed Bandit (MAB)

Dinámica Entorno	Fija	
Disponib Info	Alta	Baja

2.2 Gestion Inventarios / Inventario Rápido

Para optimizar gestion del inventario se produce en cantidades limitadas, se testea el producto en tiendas y se usa la reacción de los clientes para tomar decision sobre inventario

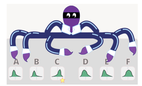
Método ágil y permite decisiones basadas en datos

Reto
El minorista se enfrenta a una situación en la que tiene una gran cantidad de productos que debe vender. Debe calcular cuántos productos se van a vender y almacenarlos antes de distribuirlos. Esto supone un gran riesgo, porque supone exponerse a grandes pérdidas. Además, esta decisión es particularmente compleja en cuanto que se dispone de muy pocos datos antes de la inversión.

Importancia
Debido a esto, la rotación de inventario se utiliza como indicador de rendimiento del negocio. Esta se define como la cantidad de veces que el inventario financiero (stock) se renueva en un periodo de tiempo. Depende de la demanda de los clientes por los productos, y de la capacidad logística de reponer ese inventario para continuar vendiendo productos. En una situación óptima, el inventario se renovará rápidamente. Es por ello fundamental que su gestión sea impecable y se actualice frecuentemente.
En el siguiente ejemplo se da por hecho que la compañía tiene una alta tasa de rotación de inventario, respaldada por fuertes despliegues comerciales y logísticos.

Caso 2: Multi-Armed Bandit

El **multi-armed bandit** es un algoritmo de refuerzo que permite tanto explorar como explotar un conjunto de opciones alternativas. El algoritmo recorre todas las opciones para evaluarlas y escoge la que proporcione una mayor recompensa. Su nombre tiene origen en la imagen de un jugador de máquinas tragaperras.



Como se determina que productos comprar y cuantos ?

- **Para los minoristas, una de las inversiones más importantes y habituales es el inventario**
- **El gran número de productos a vender es el desafío principal de la gestión de inventario**

Las decisiones de **inventario** tienen un **gran impacto** en el negocio

El concepto de **rotación de inventario** o renovación es un indicador **de rendimiento empresarial**

Desafío principal de la toma de decisiones sobre el inventario
Antes de la inversión, hay poca información disponible sobre el producto

Es una solución creativa y pragmática para los desafíos de la TDD sobre inventario

Zara, la marca de Inditex, es pionera en inventario rápido
ZARA

El motor de **toma de decisiones** puede ayudar a tomar decisiones eficaces **sobre el inventario**

Dinámica de entorno	Fija	
Disponibilidad de información	Alta	Baja

Solución: multi-armed bandit

- El entorno no cambia
  - Como los productos se introducen durante un periodo corto de tiempo, los clientes no cambian
- Disponibilidad de recopilación de datos baja
  - Cuando se introducen los productos en un periodo corto de tiempo, la recopilación de datos es relativamente baja

- Explorar con un número reducido de unidades del producto y explotar los productos que se vendan mejor
- Equilibrar exploración y explotación
- Usar el machine learning y tu conocimiento del sector para determinar cómo explorar y explotar

Multi-armed bandit contextual
- **Seleccionar productos para probarlos de tal forma que todas las características y atributos se presenten razonablemente**
- **Extrapolar los resultados a los productos no probados**

¿Cuál es el objetivo comercial que el proceso de toma de decisiones intenta optimizar?
Selecciona las respuestas correctas y haz clic en "Enviar".
- Rotación de inventario.
- Inversión máxima.
- Productos en los que invertir.

3) Proceso Decision Markov (MDP)

Dinámica Entorno	Alta	Baja
Disponib Info	Alta	Baja

2.3 Sugerencias Tiempo Real Retail

Sistema Recomend. Amazon

Amazon ha prosperado gracias a un complejo sistema de recomendación. Se recogen los datos de compra de cada cliente y se utilizan para predecir qué productos pueden interesarle. Gracias a esto, cada vez que el usuario realiza una compra, se le sugieren otros productos, incentivando así más compras.

Usuario interactúa continuamente con sistema y sistema se nutre de historial de compras

Los sistemas de recomendaciones ayudan a los usuarios a **encontrar lo que buscan** antes incluso de que pregunten

Los **sistemas de recomendaciones** ayudan a salvar la distancia entre lo que el usuario cree que ofrece un sitio web y lo que realmente ofrece.

NETFLIX YouTube amazon Etsy ebay Quora Google

Desafío: un sitio web tiene demasiados productos como para mostrarlos todos al usuario

Objetivo: mostrar solo unos pocos productos de las muchas opciones disponibles

Objetivo final: maximizar el engagement, beneficio, ingresos y/o acceso a la información

Estrategia de toma de decisiones
Construir un sistema de recomendaciones con datos como:
- Historial de cliente
- Atributos de la marca
- Elección del producto
- Información en tiempo real

Caso 3: MDP Markov

¿Qué mostrar de entre todas las opciones?

Consumo de contenido o la compra del producto

- El entorno es el historial de interacción del individuo y los atributos
- Cambia con cada interacción

El **proceso de decisión de Markov** se puede utilizar para diseñar sistemas de recomendaciones en **tiempo real**

Dinámica de entorno	Alta	Baja
Disponibilidad de información	Alta	Baja

Diseñar sistemas de recomendaciones es como jugar al ajedrez con un oponente conocido:
- Gran cantidad de datos
- El entorno cambia constantemente

Decisiones offline
- Usar el método de **iteración de política** para determinar la política óptima

Decisión en tiempo real (cuando un cliente interactúa con el sitio web)
- Usar la mejor decisión del método de iteración de política
- Reaccionar al entorno existente y proporcionar las mejores recomendaciones

Asocia cada concepto con su materialización en el contexto de los sistemas de recomendación.
Arrastra cada concepto a su grupo correspondiente y haz clic en "Enviar".

Equilibrar exploración y explotación da lugar a mejores decisiones

4) Aprendizaje x Refuerzo (AxR)

Dinámica Entorno	Alta	Baja
Disponib Info	Alta	Baja

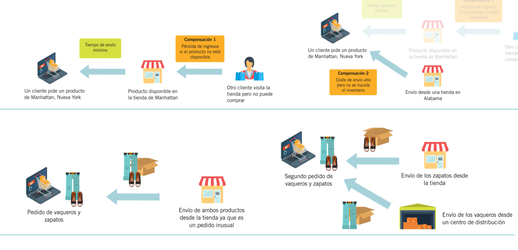
2.4 Gestion Pedidos

Decisión a que centro de distribución enviar los productos y cuando

La dificultad de esta toma de decisiones radica en la cantidad de variables que debemos tener en cuenta. Enviar a la tienda un producto demasiado pronto no es siempre viable, mientras que enviarlo demasiado tarde supondrá perder compras. Asimismo, se debe tener en cuenta que puede compensar enviar ciertos productos juntos. Carecemos de información *a priori* que nos permita tomar estar decisiones.

En un mercado online, los minoristas toman muchas decisiones sobre la gestión de pedidos:
- ¿Desde qué centro de distribución o tienda distribuyo?
- ¿Cuándo?

Objetivos de la gestión de pedidos:
- Minimizar el tiempo de envío
- Minimizar el coste de envío
- Maximizar los ingresos y ganancias



La gestión de pedidos es similar a jugar al ajedrez con un oponente desconocido

El aprendizaje por refuerzo se puede usar para la toma de decisiones sobre la gestión de pedidos, que es como jugar al ajedrez con un jugador desconocido

Aprendizaje por refuerzo para la gestión de pedidos
- Cada decisión se convierte en recompensa
- Con el tiempo, se genera mayor aprendizaje sobre el sistema y la política y da lugar a mejores decisiones

¿Cuál variables deberían configurarse como recompensas para el sistema de toma de decisiones?
Selecciona las respuestas correctas y haz clic en "Enviar".
- El tiempo de envío.
- El estado del inventario.
- El coste de envío.

3 Apps Financieras

Aprendizaje transferido

Aprendizaje transferido
Recopilar información para un objetivo y transferirla para tomar mejores decisiones para otros objetivos

- Durante la exploración, tu objetivo es recopilar información para tomar decisiones futuras
- Esta información se puede usar a un objetivo cambio

Exploración
- Usar todas las rutas para obtener información y tomar una decisión

Explotación
- Seleccionar una ruta antes de explorarlas todas

Equilibrar exploración y explotación

Recopilar información sobre las rutas

Seleccionar la mejor ruta para el objetivo

Apps Google Maps / Waze => TDD: escoger mejor ruta de un punto a otro

Inputs
- Puntos de partida y de llegada
- Información histórica y actual sobre el tráfico

Objetivo
- Minimizar el tiempo medio de viaje y su variación

Entorno
- Estado actual de los carreteros

Aplicar la política de iteración en datos hist/ricos → Entrenar la política → Aplicar la política de iteración al objetivo → Obtener la mejor ruta

3) Proceso Decision Markov (MDP)

Dinámica Entorno	Alta	Baja
Disponib Info	Alta	Baja

1 Ciudad Conocida / Markov

Selección de la mejor ruta: recopilación de datos

Inputs
- Puntos de partida y de llegada
- Rutas posibles al punto de llegada

Exploración
- Usar todas las rutas para obtener información y tomar una decisión

Explotación
- Seleccionar una ruta antes de explorarlas todas

Recopilar información sobre las rutas → Multi-armed bandit → Seleccionar la mejor ruta para el objetivo

2) Multi-Armed Bandit (MAB)

Dinámica Entorno	Fija	
Disponib Info	Alta	Baja

2 Ciudad Desconocida / MAB

Recopilar datos y usarlos mientras se recopilan

Cada ruta es un brazo del MAB

- Tasa de información muy baja
- El entorno cambia constantemente

Inicialmente se incurre en pérdidas, pero el resto se ve beneficiado

Inputs
- Estado actual de los carreteros
- Datos previos o actuales para futuras predicciones

Objetivo
- Minimizar el tiempo de viaje

Entorno
- Entorno cambiante

Ruta 160

Ejemplo: un nuevo puente que conecta con carretera

En este entorno el mundo acaba de cambiar

Cada vez que vehículo viaja transmite info

4) Aprendizaje x Refuerzo (AxR)

Dinámica Entorno	Alta	Baja
Disponib Info	Alta	Baja

3 Ciudad con Obras / AxR

4 Mejor Ruta

5 AlphaGo Zero

Desarrolló estrategia de aprendizaje

Modelo de autojuego de oponente

AlphaGo Zero es un **jugador de Go automatizado** que vence a los mejores jugadores sin **datos históricos**

Sistema TDD AlphaGo: capaz de aprender las reglas y las mejores estrategias del juego sin contar con ningún dato previo de otros jugadores, sin mayor esfuerzo computacional

Para encontrar la política óptima:
- Usar el modelo predictivo para aprender valores de cada frame
- Determinar la recompensa esperada para el siguiente movimiento
- Forzar la política para que se mueva hacia el siguiente estado con mayor recompensa
- Seleccionar la secuencia de movimientos que maximiza la recompensa final

Challenge: Este método requiere una gran cantidad de datos históricos

Para encontrar la política óptima:
- Usar el modelo predictivo para aprender valores de cada frame
- Determinar la recompensa esperada para el siguiente movimiento
- Forzar la política para que se mueva hacia el siguiente estado con mayor recompensa
- Seleccionar la secuencia de movimientos que maximiza la recompensa final

Para encontrar la política óptima:
- Usar el modelo predictivo para aprender valores de cada frame
- Determinar la recompensa esperada para el siguiente movimiento
- Forzar la política para que se mueva hacia el siguiente estado con mayor recompensa
- Seleccionar la secuencia de movimientos que maximiza la recompensa final

Input Estado del juego → **Política** Red neuronal con dos cabezas → **Output** Movimiento óptimo

La simulación de partidas genera gran cantidad de datos de entrenamiento para diferentes posiciones del tablero y predice la recompensa final del juego

Comenzar con una política → Simular un número de partidas de entrenamiento con la política → Revisar movimientos ejecutados en cada partida → Perfeccionar la política → Ajustar una función de pérdida para que la predicción de ganar o perder se ajuste a los datos

Determinar la probabilidad de ganar o perder

El aprendizaje supervisado se consigue a través de redes neuronales ya que se trata de un problema de clasificación

Los datos de entrenamiento que recaba y utiliza AlphaGo Zero asocian:
Selecciona la respuesta correcta y haz clic en "Enviar".
- Movimientos específicos con las reacciones del rival.
- Combinaciones de movimientos con las reacciones del rival y los resultados de victoria o derrota.
- Movimientos específicos con resultados de victoria o derrota.