

Repaso de técnicas

Programa en Analítica

Curso *Capstone* - Inteligencia Artificial & *Deep Learning*

Analítica Prescriptiva

Educación continua | Universidad de los Andes
octubre 6 – noviembre 24
2021

Hoy

De qué vamos a hablar

1. Proyecto

¿Qué nos vamos a llevar de esta clase?

2. Desfile de técnicas

¿Qué tipo de situaciones ataca? ¿Cómo se concluye con? ¿De qué datos se parte?

3. Siguiendo paso:

*Tareas para elaborar durante siguiente clase.
Objetivos de consulta.*



Proyecto

Qué nos llevamos hoy

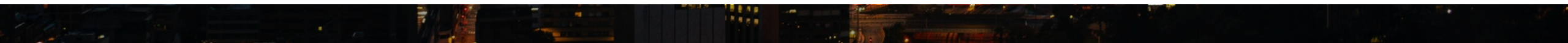


Photo by [nappy](#) from [Pexels](#)

En aprendizaje basado en problemas

... queremos formar autonomía.

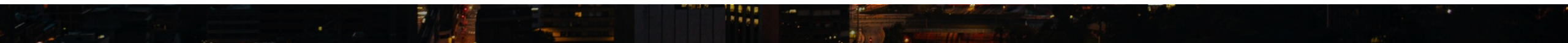
- Por eso el profesor es un facilitador que desbloquea, redirecciona, etc.
- Pero cada estudiante explora y se va encontrando con retos e información nueva.



En aprendizaje basado en problemas

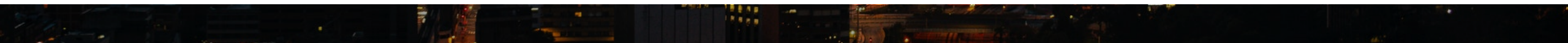
... queremos formar autonomía.

- Por eso el profesor es un facilitador que desbloquea, redirecciona, etc.
- Pero cada estudiante explora y se va encontrando con retos e información nueva.
- Es activo, es práctico y genera aprendizaje significativo.



En aprendizaje basado en problemas

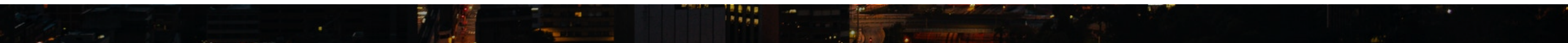
Ya hemos pasado por



En aprendizaje basado en problemas

Ya hemos pasado por

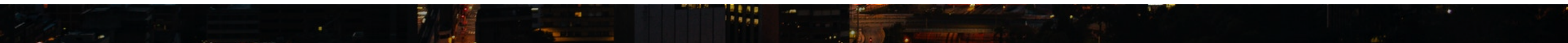
1. Definición de palabras y términos con los que describimos el problema.



En aprendizaje basado en problemas

Ya hemos pasado por

1. Definición de palabras y términos con los que describimos el problema.
2. Definición y formalización del problema.

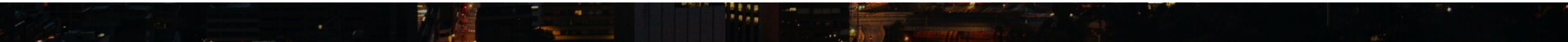


En aprendizaje basado en problemas

Ya hemos pasado por

1. Definición de palabras y términos con los que describimos el problema.
2. Definición y formalización del problema.

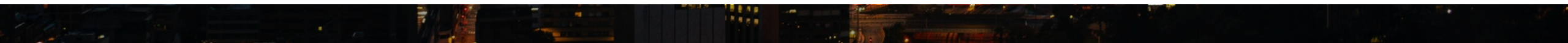
Identificamos y definimos el problema



En aprendizaje basado en problemas

Ya hemos pasado por

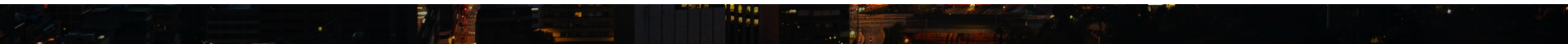
1. Definición de palabras y términos con los que describimos el problema.
2. Definición y formalización del problema.
3. *Brainstorm* de aproximaciones:



En aprendizaje basado en problemas

Ya hemos pasado por

1. Definición de palabras y términos con los que describimos el problema.
2. Definición y formalización del problema.
3. *Brainstorm* de aproximaciones:
 - Ya hemos explorado algunas ideas de estrategia general
 - Hoy vamos a volver sobre técnicas que hemos mencionado en el programa



En aprendizaje basado en problemas

Ya hemos pasado por

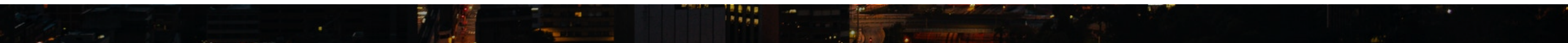
1. Definición de palabras y términos con los que describimos el problema.
2. Definición y formalización del problema.
3. *Brainstorm* de aproximaciones:
 - Ya hemos explorado algunas ideas de estrategia general
 - Hoy vamos a volver sobre técnicas que hemos mencionado en el programa

VAMOS ANOTANDO QUÉ PODRÍA SERVIRNOS

De aquí pueden salir preguntas

1. Técnicas: sobre cómo aplicar / programar, etc.
2. Teóricas: para entender bien qué hace y cómo lo hace la técnica
3. Otras ...

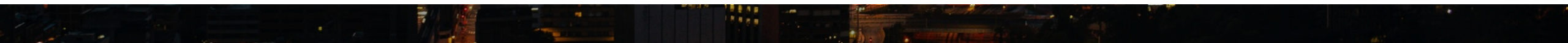
Vamos anotándolas.



¿Qué haremos al final de la clase?

1. Compartir nuestros apuntes en los grupos:

- Tareas concretas de cosas que vamos a intentar
- **Objetivos de consulta:** preguntas que cada uno va a buscar ahorita o la otra clase. Ej. Qué **paquete** se usa en R para correr _____ técnica. Ej. Buscar un **tutorial** de ejemplo de uso de esta técnica.
- Nos asignamos tareas para aprovechar nuestro equipo la siguiente clase



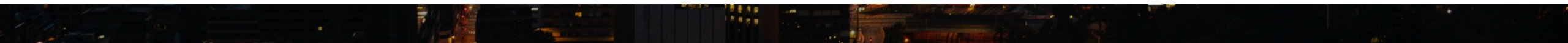
¿Qué haremos al final de la clase?

1. Compartir nuestros apuntes en los grupos:

- Tareas concretas de cosas que vamos a intentar
- **Objetivos de consulta:** preguntas que cada uno va a buscar ahorita o la otra clase. Ej. Qué **paquete** se usa en R para correr _____ técnica. Ej. Buscar un **tutorial** de ejemplo de uso de esta técnica.
- Nos asignamos tareas para aprovechar nuestro equipo la siguiente clase

2. La siguiente clase:

- Reportamos exploraciones individuales en los grupos
- Intentamos estimar y correr modelos



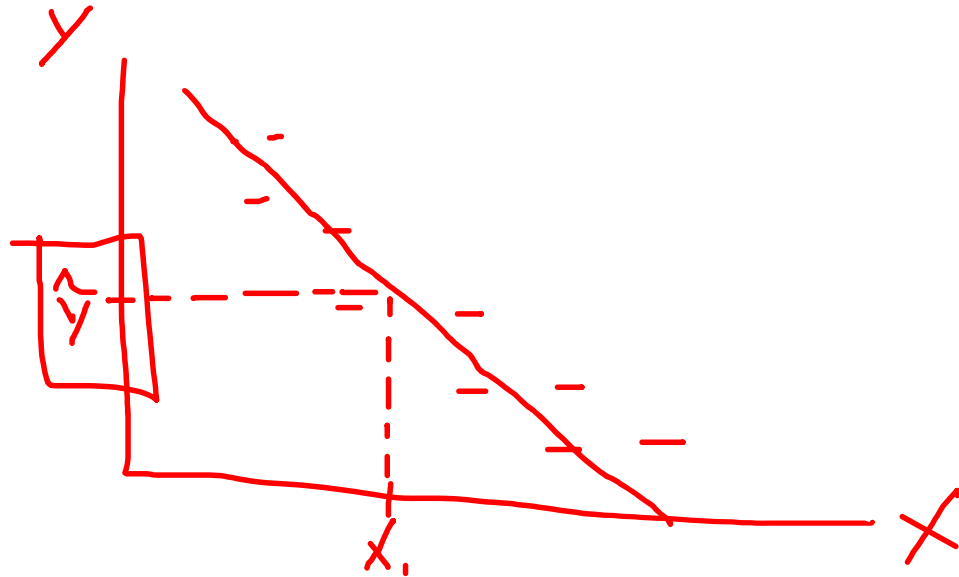
Desfile de técnicas

Un repaso a grandes rasgos



Photo by [nappy](#) from [Pexels](#)

Regresión lineal



$$Y = \beta_0 + \beta_1 X$$

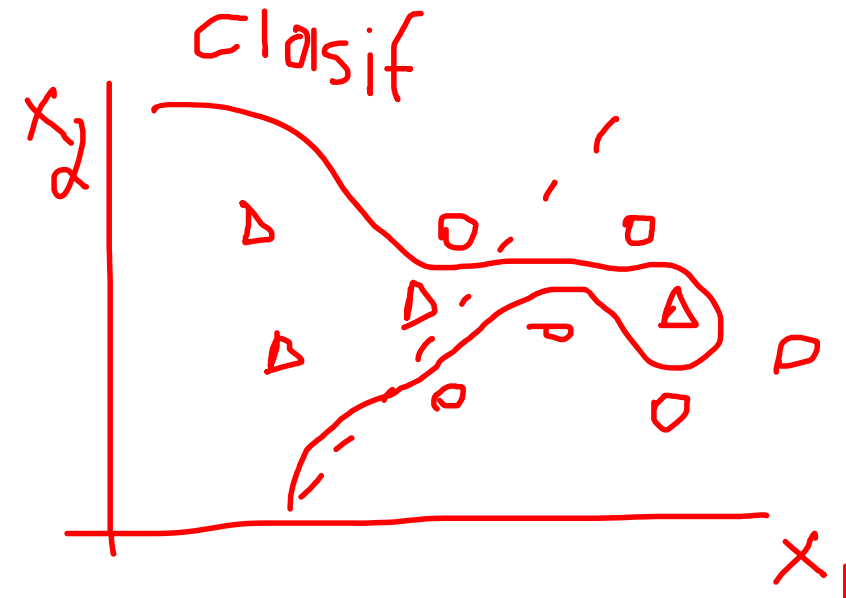
- Ajustamos una línea a la relación entre características y variable etiqueta, graficados en un espacio.
- Predecimos para ciertas características, un valor de etiqueta.
- Conclusión: este dato nuevo, tiene tales características, se espera que tenga este valor de etiqueta.

Regresión polinomial

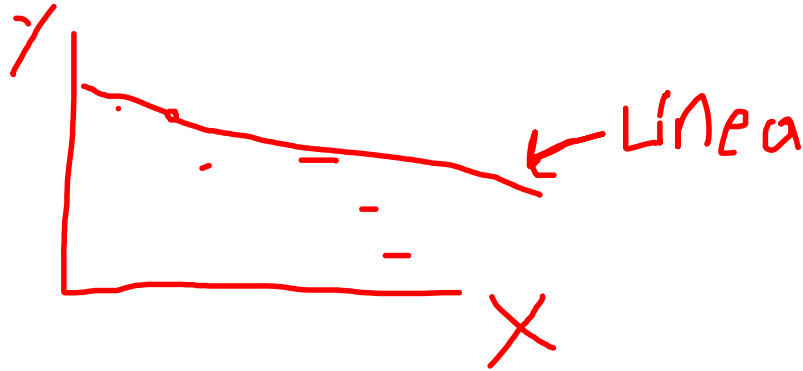
- Agregamos más características al modelo, e incluimos grados superiores a 1.
- Conclusión: tal observación con tales características, tiene un valor de etiqueta tal.
- Más complejo no necesariamente predice mejor.. Pero... cuando vamos a clasificar: podemos incrementar detalle.

y	x	x^2

$$y = \beta_0 + \beta_1 x + \beta_2 x^2$$



Técnicas de regularización (L1, L2 y elástica)

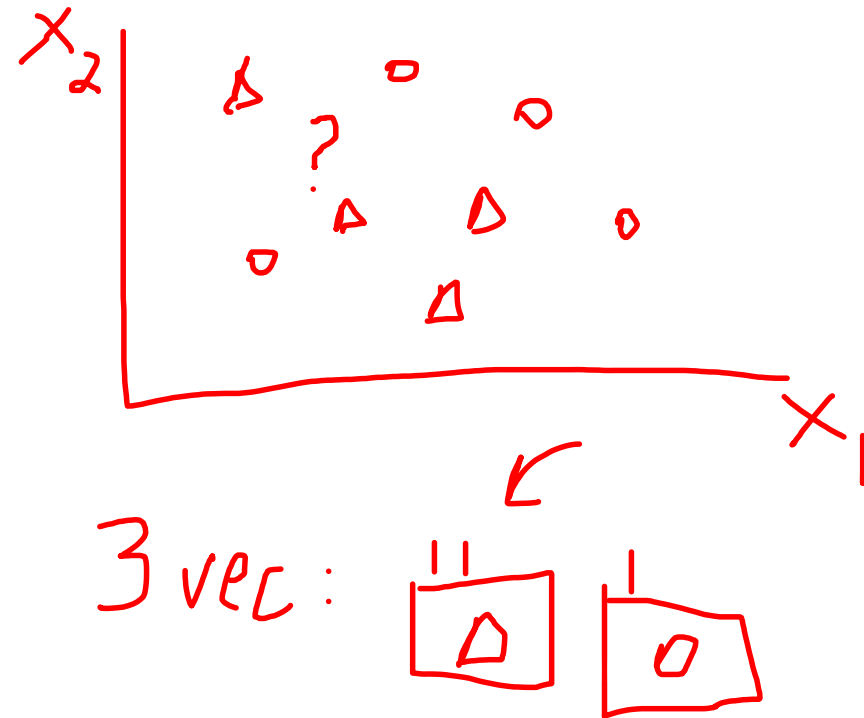


- Ajuste ✓
- Rigidez

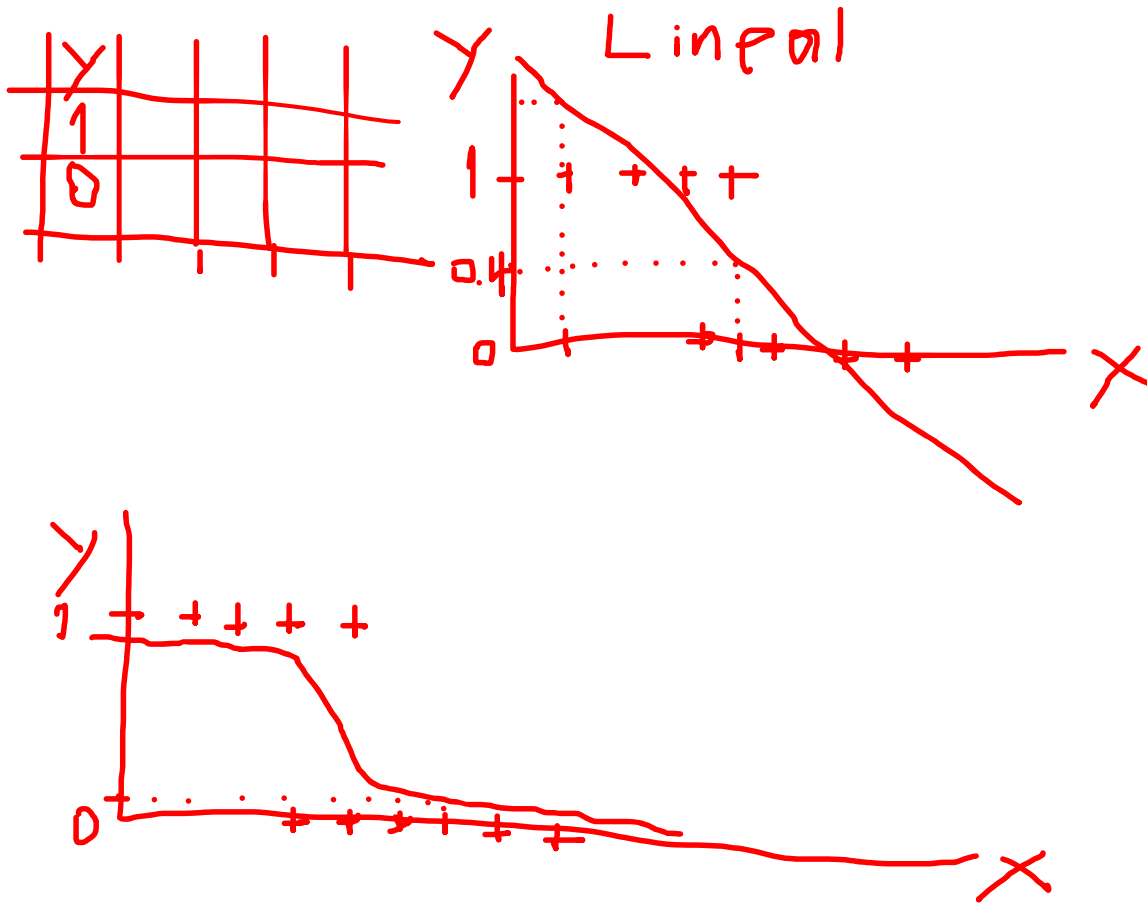
- Ajustar una línea significa que ajustamos su pendiente y su "altura" o intercepto.
- Lo hacemos de forma tal que ajusta lo mejor posible los datos.
- Para mitigar efectos de sobreajuste, podemos imponer rigideces a la línea: no queda tan ajustada pero puede generalizar mejor.

KNN – K vecinos más cercanos

- Clasificamos: graficamos la nueva observación en un espacio de características y votamos los k vecinos más cercanos según su tipo.
- Ajustamos el K, probamos con k's impares para evitar empates.



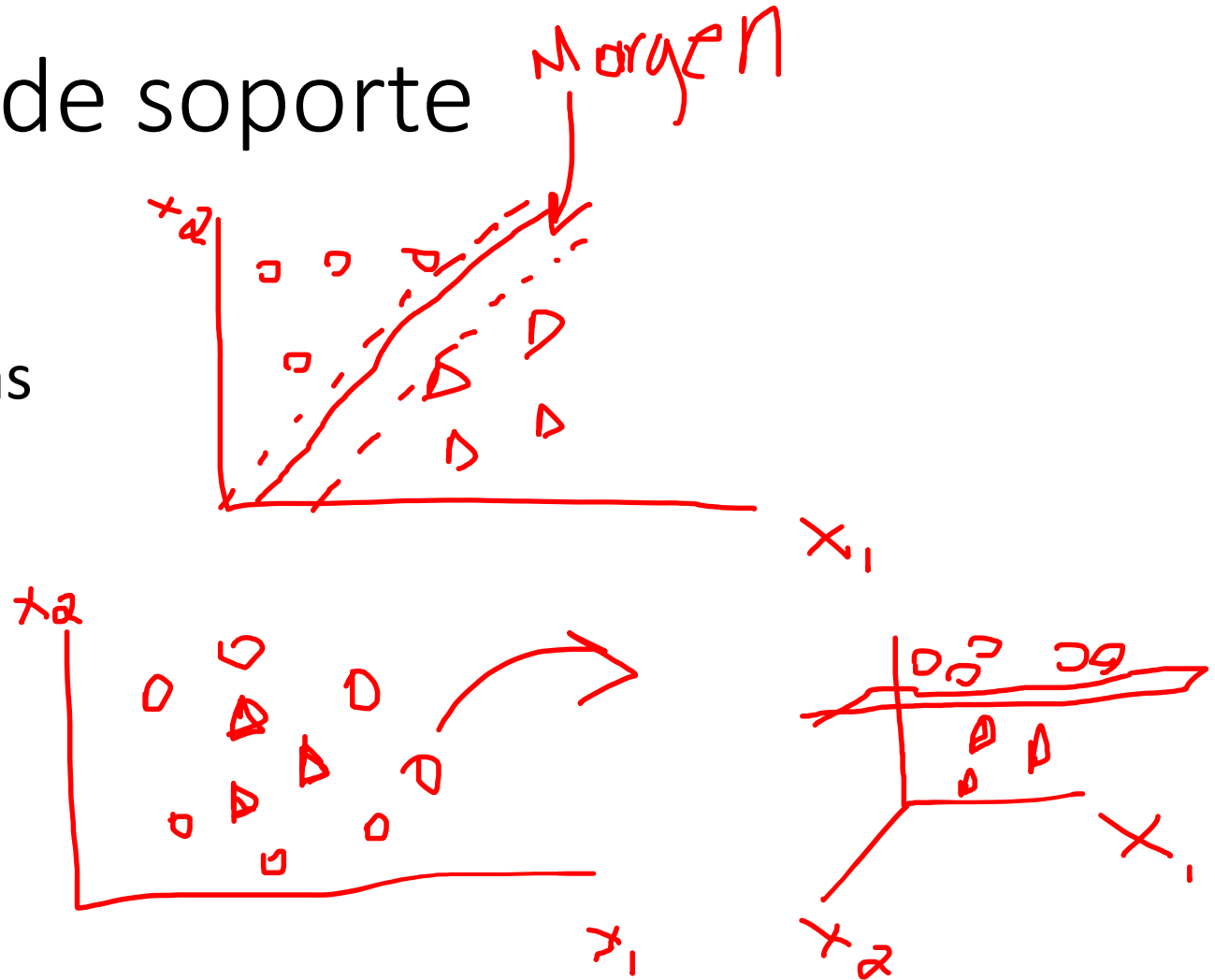
Clasificación con modelo logístico



- Modelo lineal nos permite clasificar prediciendo la probabilidad de una categoría.
- Lineal nos da probabilidades mayores a 1 y menores a 0.
- Podemos usar una función asintótica en 1 y 0.
- ¿Cambia mucho el desempeño? Hay que probar.

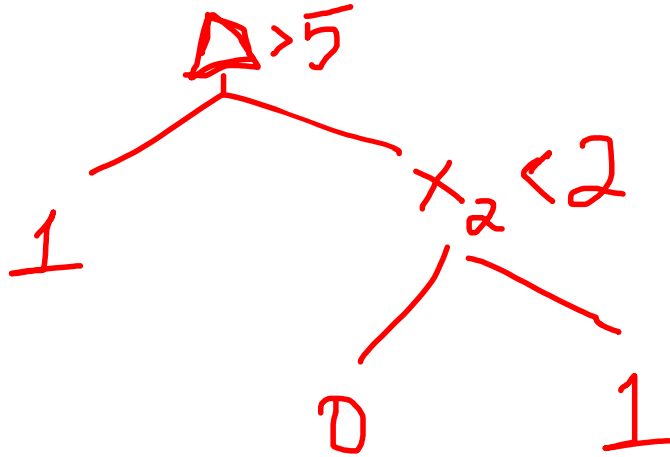
Máquina de vectores de soporte

1. La máquina escoge un (hiper) plano que separa las categorías en un espacio maximizando el margen entre categorías.
2. Si no se pueden separar... se puede aplicar el truco del Kernel.



Volvemos a las 7:12

Árboles de decisión

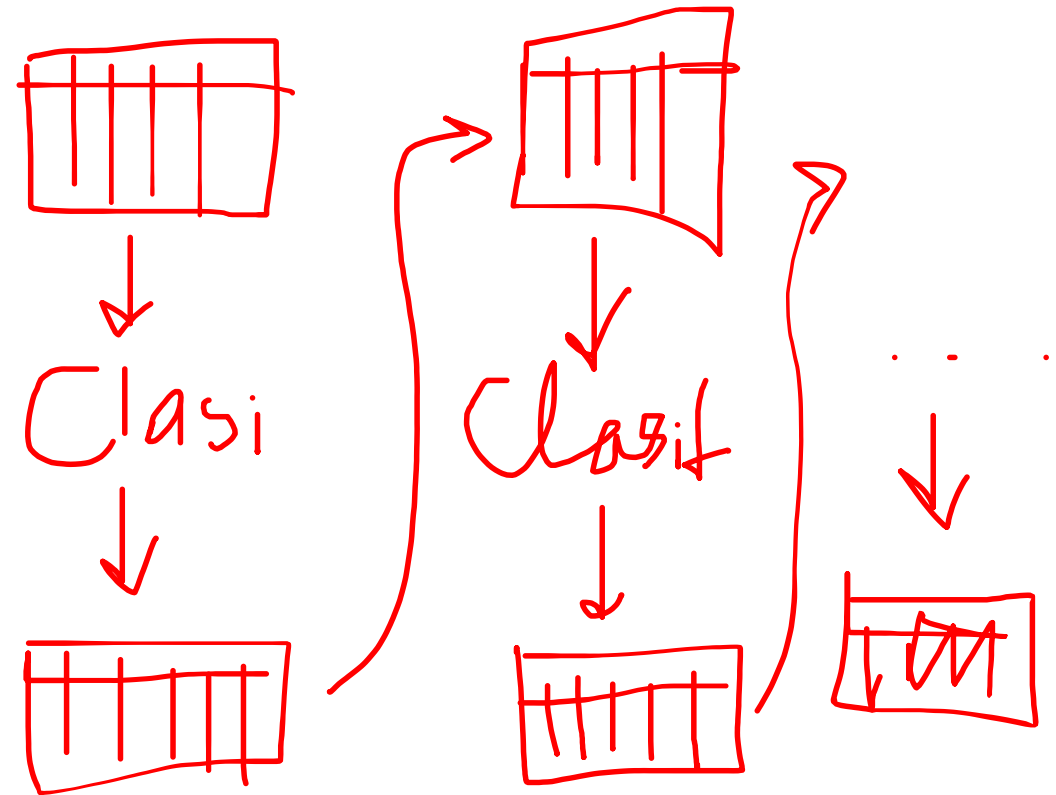


x_1	x_2	x_3	

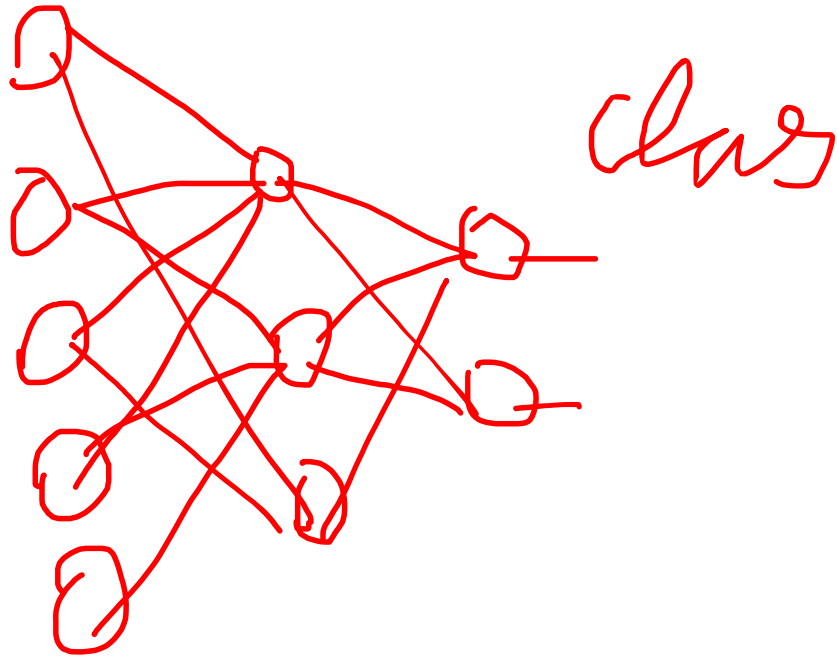
- Partimos de la característica con mayor correlación con la etiqueta a predecir.
- Hacemos una pregunta.
- Luego tomamos la segunda característica más correlacionada.
- Hacemos otra pregunta...
- Hasta llegar a las hojas.

Boosting

- Es una forma de tomar clasificadores débiles y volverlos uno fuerte.
- Tomamos los datos, y los pasamos por un clasificador débil.
- Ponderamos los datos y los pasamos por un segundo clasificador débil.
- Hasta la predicción final.



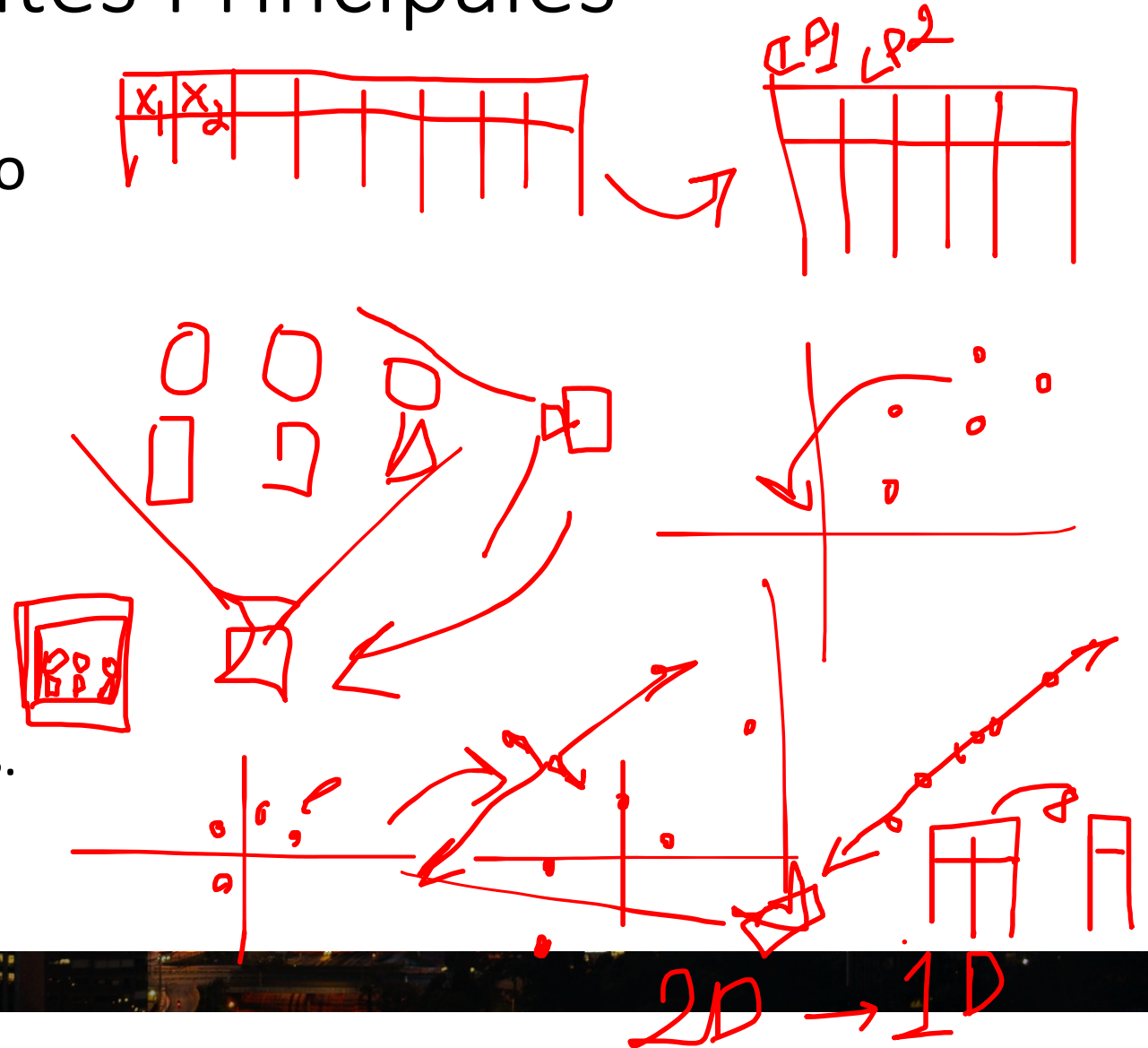
Redes neuronales



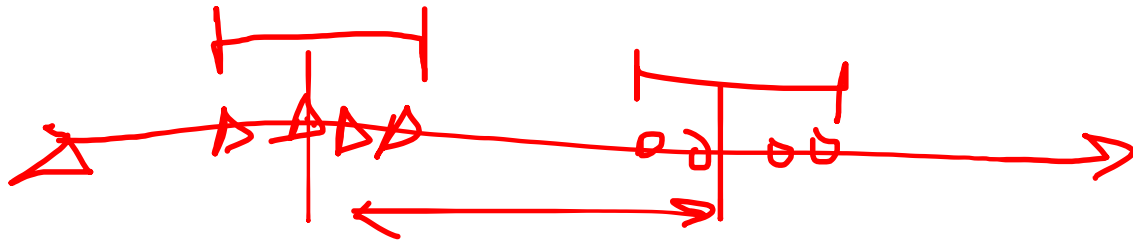
- Nos sirven para clasificar y predecir valores (regresión y clasificación).
- Tomamos una características y vamos modificando los datos capa por capa hasta...
- Predecir unas categorías o valores finales.
- Variamos la arquitectura.

Análisis de Componentes Principales

1. Queremos tomar una foto pero no todo el grupo se ve.
2. ACP nos indica el ángulo para tomar la foto:
 1. Centramos y normalizamos los datos.
 2. Estiramos el espacio en una dirección y en una magnitud.
 3. Tomamos la foto: generamos nuevas columnas interpretables.
3. Perdemos la menor info. posible.



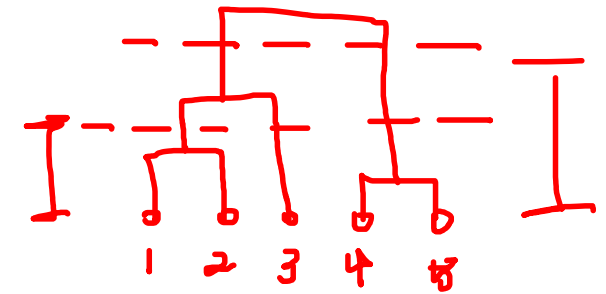
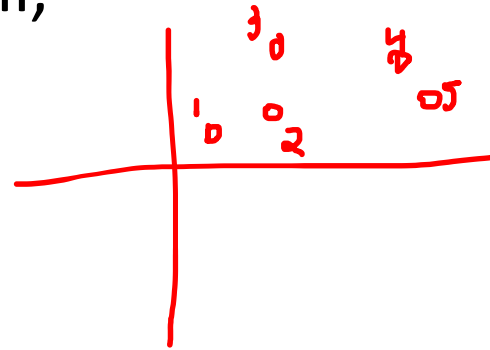
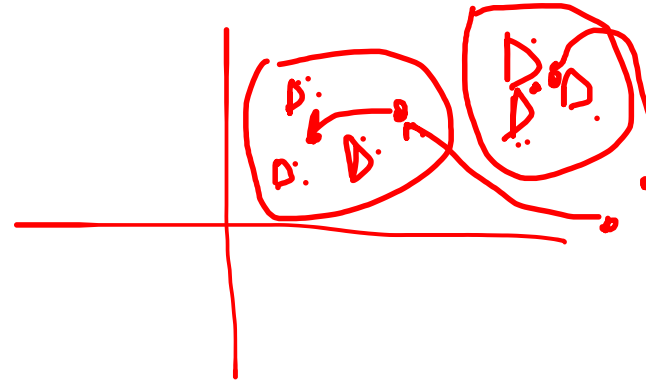
Análisis de discriminante lineal



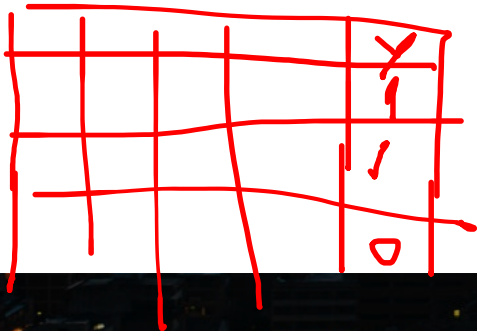
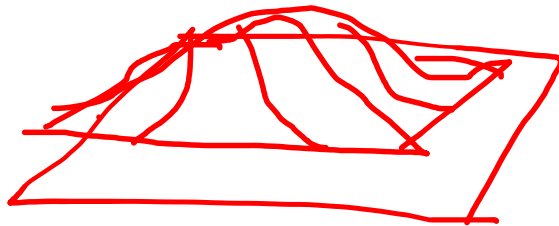
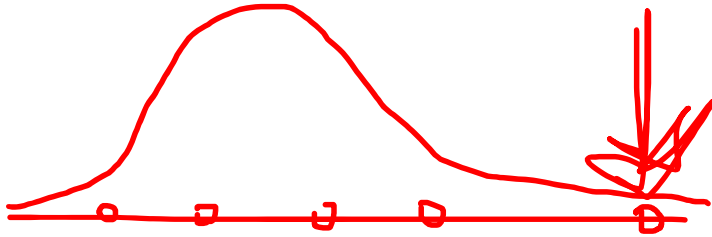
- ACP siempre genera componentes lineales. ADL está pensado para facilitar clasificación.
- Escoge dimensiones que maximizan la varianza entre las medias de las clases (proyectadas)
- Pero además, minimizan la varianza dentro de cada clase (proyectada)

Clustering (k-medias y jerárquicos)

- No supervisado.
- Tenemos datos que queremos agrupar de alguna forma.
- Para un humano puede ser fácil, para un computador no.
- Calculamos las distancias y centramos iterativamente.. O
- Construimos grupos jerárquicos.



Detección de anomalías



- Identificamos observaciones que son atípicas respecto a las demás.
- Existen dos tipos de detección de anomalías:
 1. No supervisado: asumimos que los datos son normales
 2. Supervisado: tenemos datos etiquetados como “normales” o “anormales”

Siguiente paso

Comentario



Photo by [nappy](#) from [Pexels](#)

Ahora...

1. Compartir nuestros apuntes en los grupos:

- ✓ • Tareas concretas de cosas que vamos a intentar
- Objetivos de consulta: preguntas que cada uno va a buscar ahorita o la otra clase. Ej. Qué **paquete** se usa en R para correr _____ técnica. Ej. Buscar un **tutorial** de ejemplo de uso de esta técnica.
- Nos asignamos tareas para aprovechar nuestro equipo la siguiente clase .