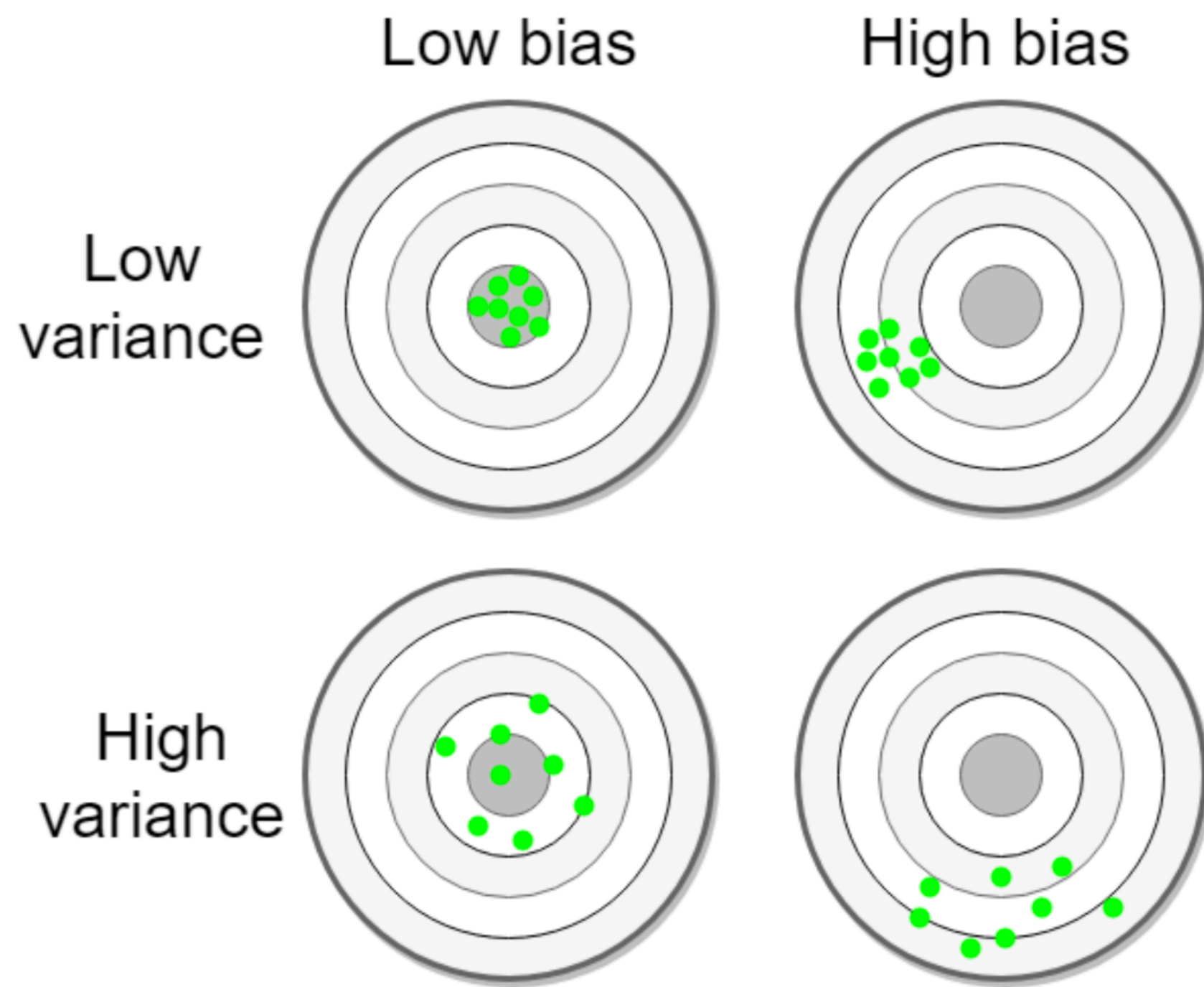
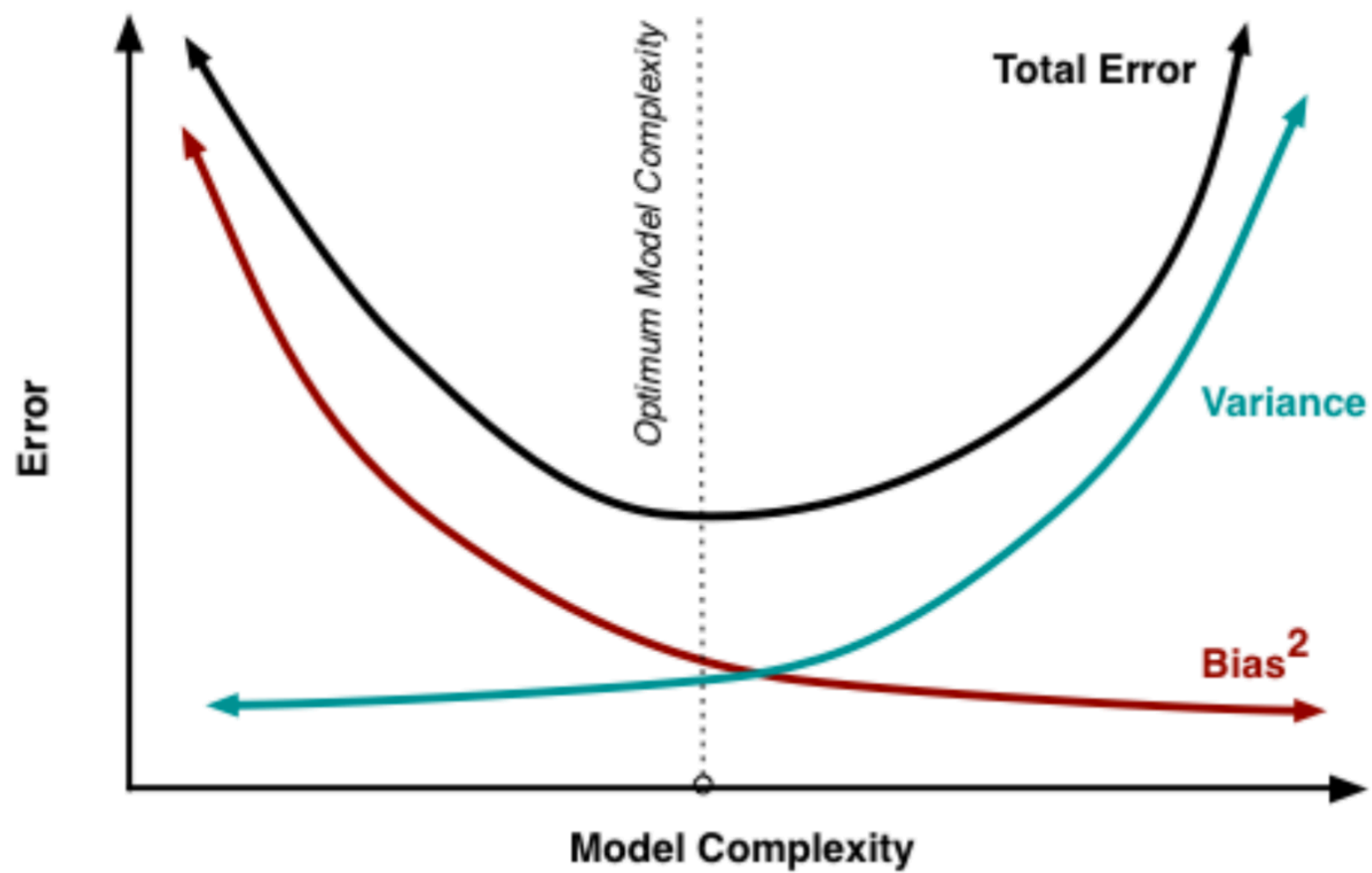


Buscamos la solución más adecuada en el **espacio de hipótesis**, usando **conocimiento previo** y datos de entrenamiento para guiar la **búsqueda**.





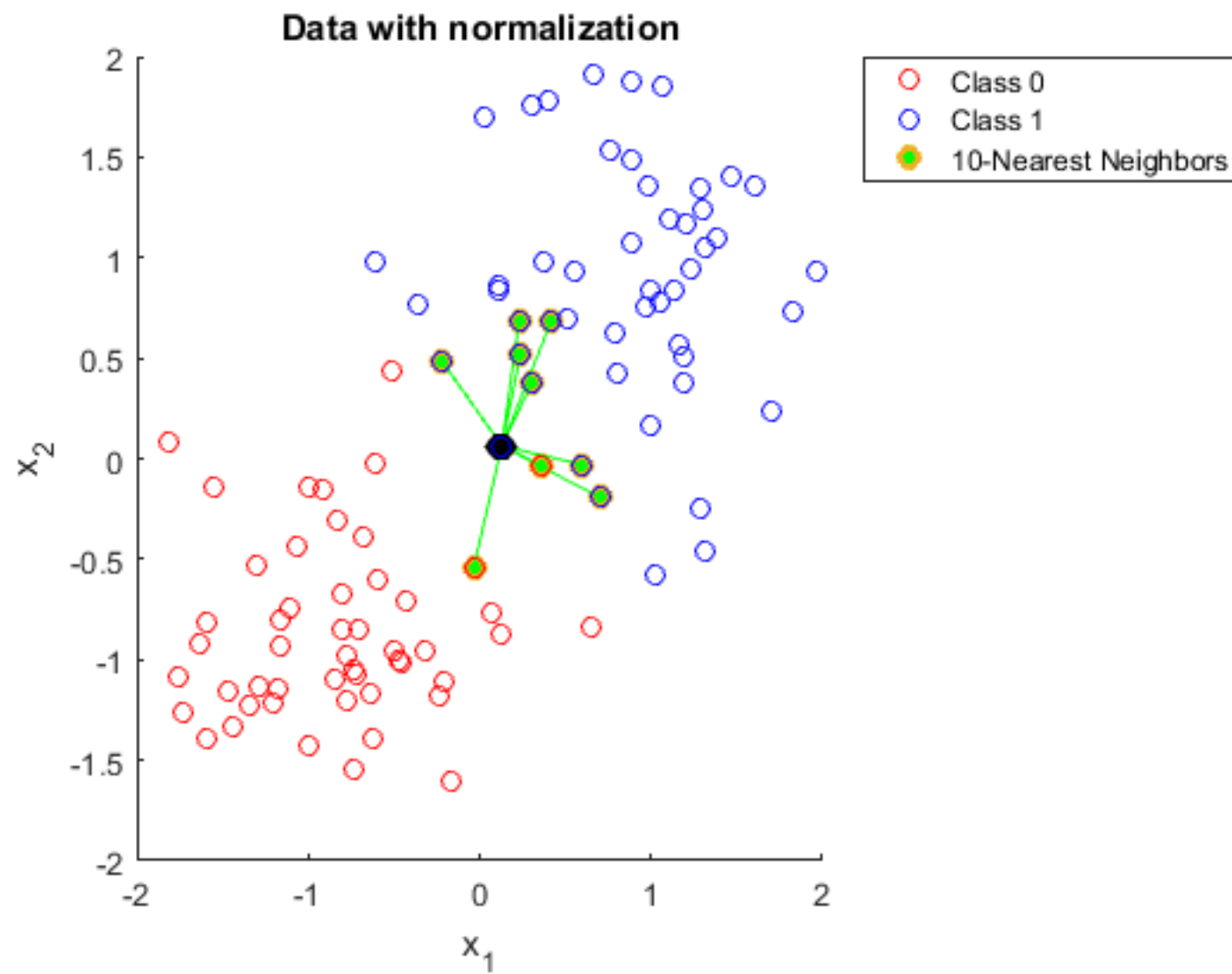
$$Err(x) = \text{Bias}^2 + \text{Variance} + \text{Irreducible Error}$$

Occam's Razor

The cyclic multiverse has multiple branes - each a universe - that collided, causing Big Bangs. The universes bounce back and pass through time, until they are pulled back together and again collide, destroying the old contents and creating them anew.

God did it.

KNN



Ejercicios

- 1.2. Según la visión genérica de un algoritmo de aprendizaje de máquina, el uso de una métrica de rendimiento permite encontrar la hipótesis que presenta la mejor capacidad de generalización medida en un set de test.

- 1.3. El problema de sobreajuste (overfitting) consiste en que el rendimiento medido en el set de test es menor al error observado en el set de entrenamiento.

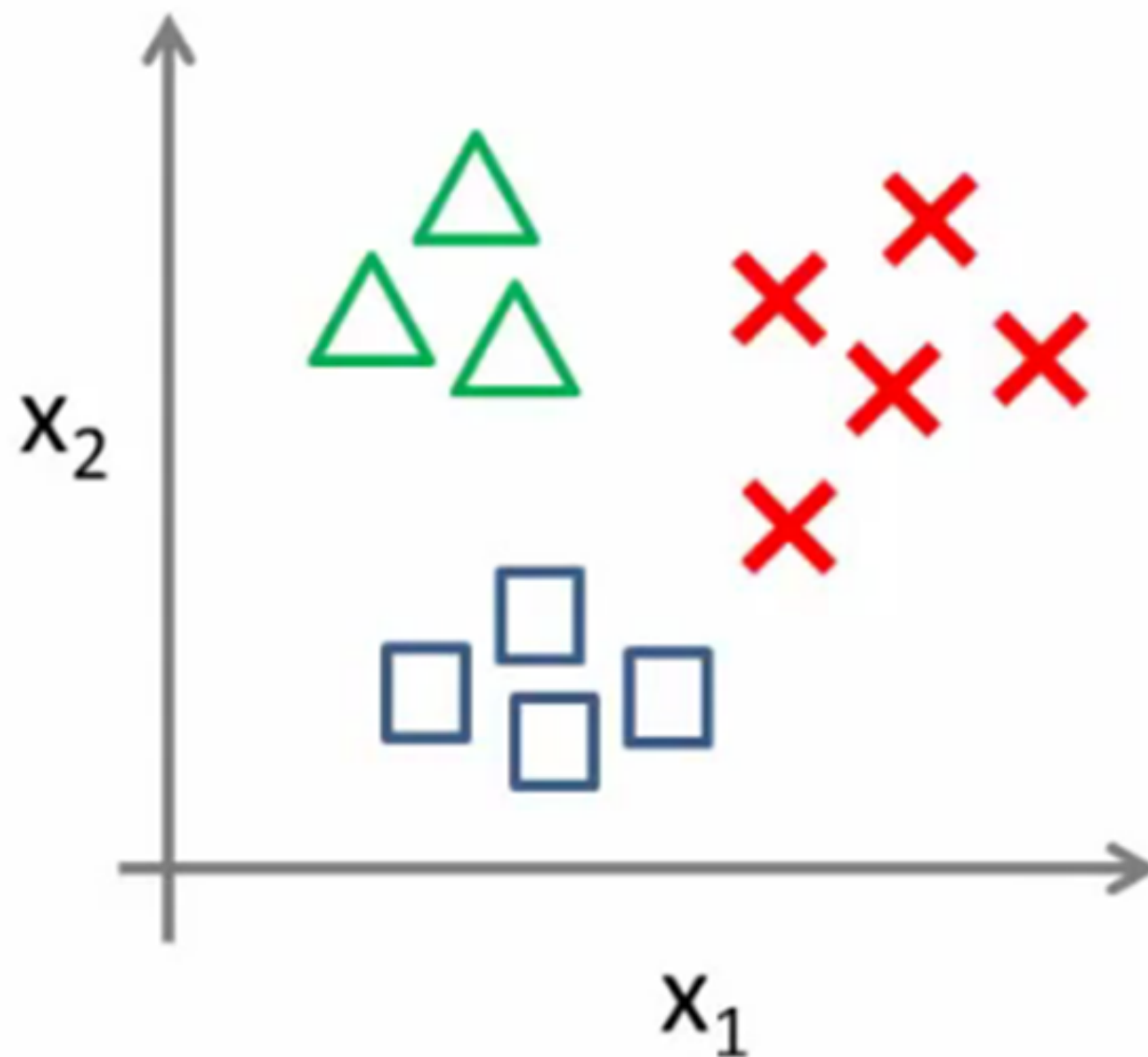
- 1.6. En general, en problemas de aprendizaje de máquina se puede asegurar que mientras más datos de entrenamiento sean usados, mejor será el rendimiento del modelo resultante en términos de sus capacidades de generalización.

- 1.4. Si un algoritmo alcanza un 100% de efectividad en el set de entrenamiento, entonces se garantiza que la hipótesis respectiva tendrá un alto grado de generalidad para clasificar nuevas instancias.

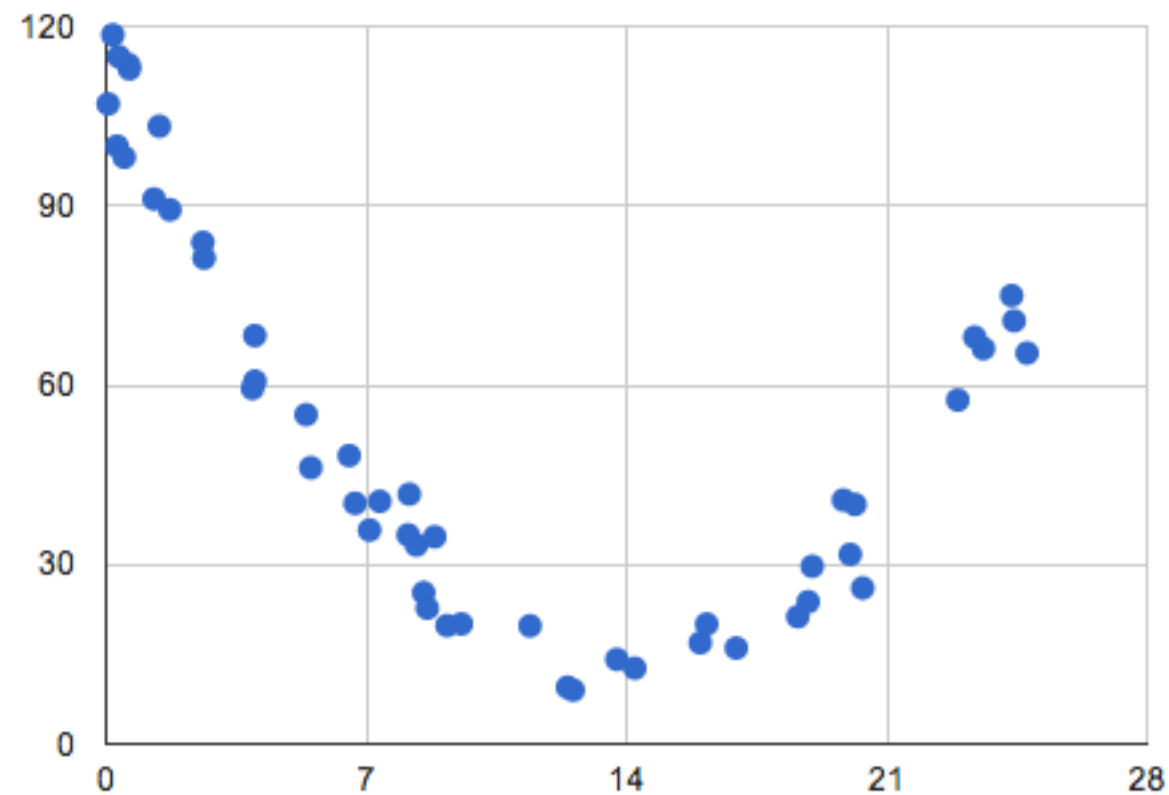
- 1.5. Mientras mayor sea la capacidad de representación de un algoritmo de aprendizaje de máquina (mayor espacio de hipótesis), menor será la probabilidad de sufrir problemas de sobreajuste.

¿Clasificador Multiclase?

Multi-class classification:



¿Relaciones no lineales en regresión?



Pregunta 1

Modifique la función de costo vista en clases para la regresión lineal

$$J(\theta) = \frac{1}{2} \sum_{i=1}^m (h(x^{(i)}) - y^{(i)})^2$$

de tal manera que dado un modelo

$$h(x) = \sum_{i=1}^n \theta_i x_i$$

con $n \in \mathcal{N}$, se pueda prevenir el overfitting sin necesidad de modificar la complejidad del modelo. Es decir sin agregar o quitar parámetros. (Hint: tenga en consideración el valor de los parámetros)

Pregunta 3

- a) Un fabricante de hardware se encuentra en el proceso de diseño de una nueva generación de CPUs para dispositivos móviles, cuya principal característica es un mecanismo adaptativo de activación/desactivación de los núcleos (cores), en base a al comportamiento de los usuarios. Con el fin de ponerse a tono con el mercado mundial, el equipo de desarrollo planea utilizar técnicas de aprendizaje de máquina para construir el sistema de activación de núcleos.
- i) Indique como podría enfrentar el equipo desarrollador este problema. Especifique claramente que tipo de aprendizaje utilizaría, y qué ventaja tendría esto por sobre un sistema basado en reglas. **(2 ptos.)**
 - ii) Con el fin de entrenar los modelos, el equipo de desarrollo recolectó un gran volumen de datos de uso de las nuevas CPUs, desde los computadores de sus propios integrantes. Comente sobre las ventajas y desventajas de esta decisión. **(2 ptos.)**
- b) Con el fin de categorizar a los alumnos de la Escuela de Ingeniería en uno de 10 posibles perfiles profesionales, un investigador decide utilizar una regresión lineal con polinomio de grado 1, sobre un conjunto de datos con distintas mediciones hechas a los estudiantes. Los registros de este conjunto se ubican en un espacio de características de 37 dimensiones, cada una con dominio en \mathbb{R} . Para calibrar los parámetros del modelo de regresión, el investigador decide utilizar un método distinto al del descenso del gradiente, debido a la posibilidad de caer en mínimos locales. Comente sobre las decisiones que tomó el investigador para llevar a cabo el estudio. **(2 ptos.)**