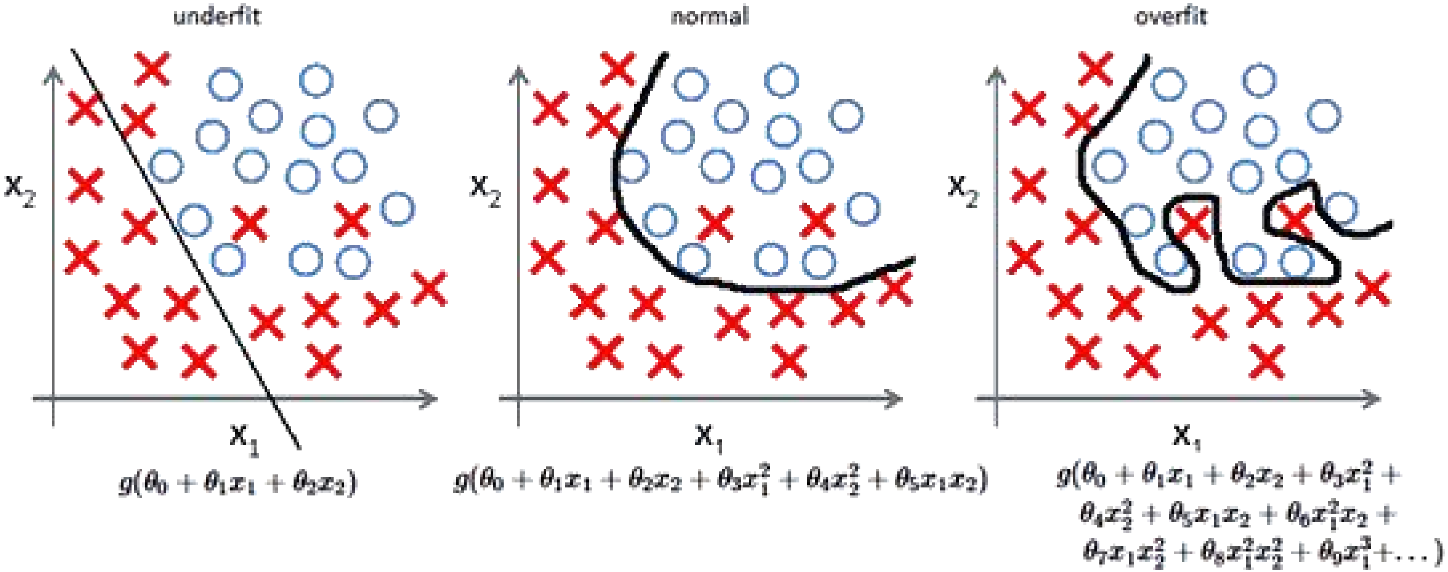
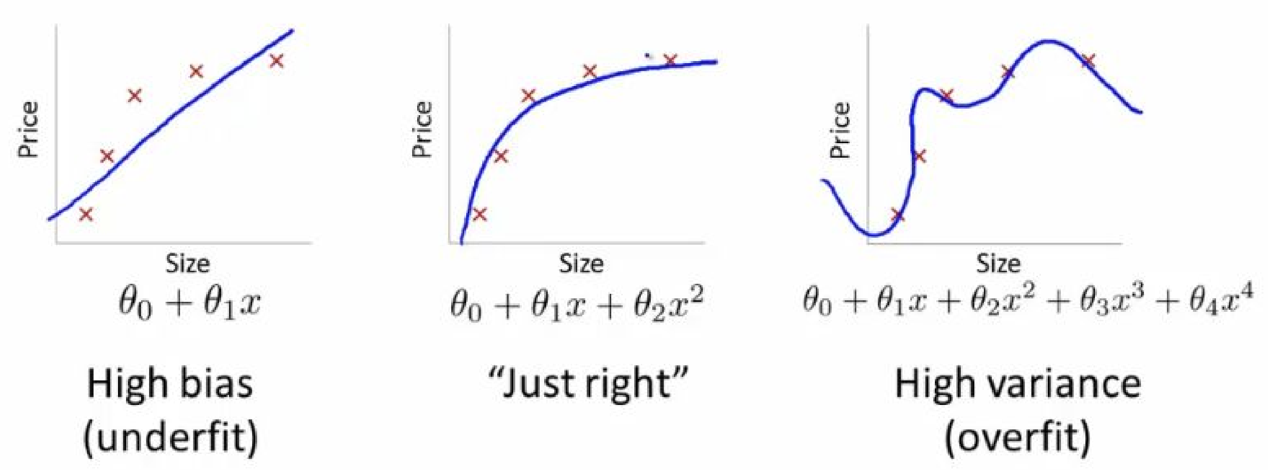
Regularización

Overfitting y Underfitting



Overfitting y Underfitting

- Bias: Error introducido por aproximar un fenómeno real con un modelo simple.
- Variance: Qué tanto cambia el error en el testing set cuando cambiamos la información de entrenamiento



Soluciones

- Reducir el número de features - pierdes información
  - Algoritmos de selección de modelo
  - Seleccionar qué features mantener

- Regularization
  - Reduce la magnitud de los parámetros.
  - Suele funcionar, en especial si todos los features importan.

Regularización

- ¿Qué pasa si penalizamos que un parámetro en la función de error sea grande?

$$\min_{\theta} \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2 + 1000\,\theta_3^2 + 1000\,\theta_4^2$$

Regularización

- Los valores de los parámetros son más pequeños
- Se hace una hipótesis más sencilla
- Menos propenso a overfitting.

Regularización para Regresión Linear con GD

$$J(\theta) = \frac{1}{2m}\left[\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2 + \lambda\sum_{j=1}^{n}\theta_j^2\right]$$

$$\min_\theta J(\theta)$$

- ¿Cómo afecta Gradient Descent?

Regularización para Regresión Linear con EN

- Antes:

$$\Theta = (X^T X)^{-1} X^T y$$

- Ahora

$$\Theta = (X^T + \lambda \begin{bmatrix} 0 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{bmatrix})^{-1} X^T y$$

- Se puede demostrar que siempre se invierte

Regularización pare Regresión Logística

$$J(\theta) = -\left[\frac{1}{m}\sum_{i=1}^{m} y^{(i)} \log h_\theta(x^{(i)} + (1 - y^{(i)}) \log (1 - h_\theta(x^{(i)}))\right]$$

$$+ \frac{\lambda}{2m} \sum_{j=1}^{n} \Theta_j^2$$

- Se suma el término de regularización
- GD es exactamente igual