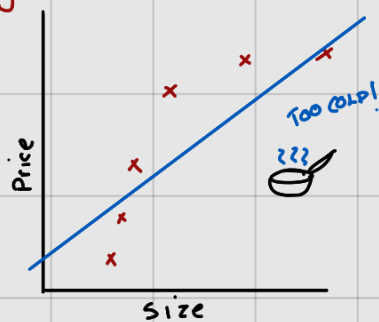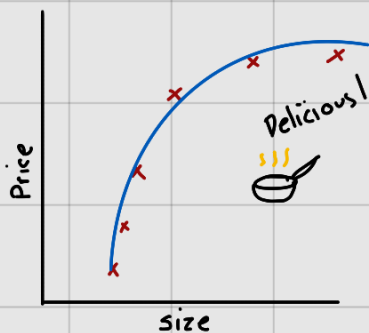# Overfitting

Regression Example

$$w_1 x + b$$

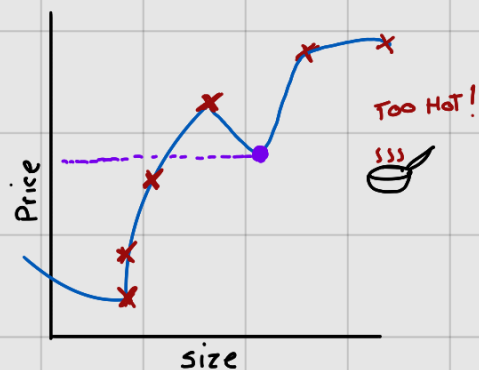**Under Fit**

- Does not fit the training set well

high bias

$$w_1 x + w_2 x^2 + b$$

Just right

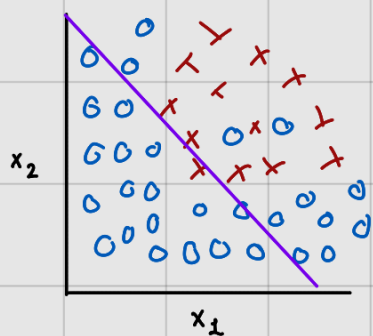- Fits training set pretty well

generalization

Too Hot!

$$w_1 x + w_2 x^2 + w_3 x^3 + w_4 x^4 + b$$

Over Fit

- Fits the training set extremely well

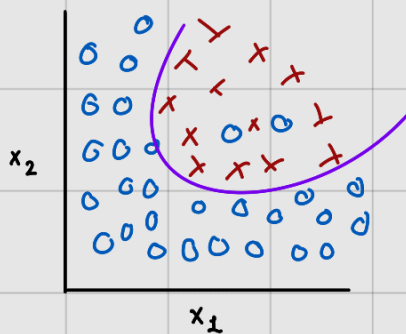**high variance**

Classification

$$z = w_1 x_1 + w_2 x_2 + b$$

$$f_{(\vec{w},b)}(\vec{x}) = g(z)$$
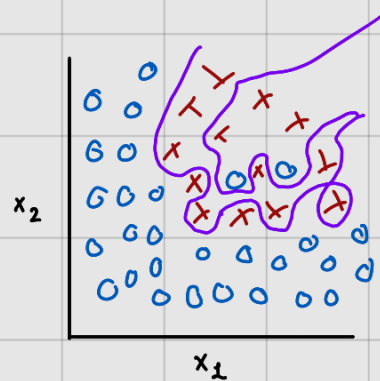
g is the sigmoid function

underfit   high bias

$$z = w_1 x_1 + w_2 x_2 + w_3 x_1^2 + w_4 x_2^2 + w_5 x_1 x_2 + b$$

Just right!

$$z = w_1 x_1 + w_2 x_2 + w_3 x_1^2 x_2 + w_4 x_1^2 x_2^2 + w_5 x_1^2 x_2^3 + w_6 x_1^3 x_2 + \dots + b$$

Overfit

Adressing over fitting

Options

1. Collect more data

2. Select features
   - Feature selection

3. Reduce size of parameters
   - "Regularization"