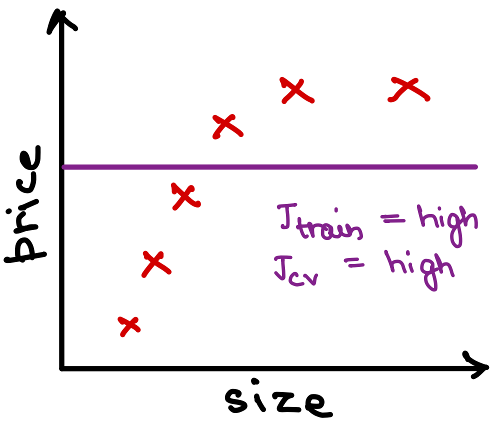
We're going to see what regularization does to bias and variance in order to understand when to use it.

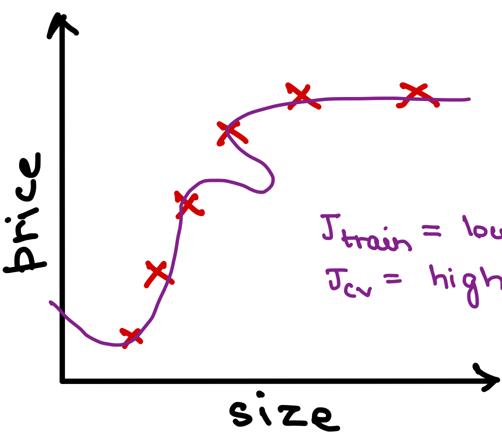
Example model: $f_{\vec{w},b}(\vec{x}) = w_1 x + w_2 x^2 + w_3 x^3 + w_4 x^4 + b$

$$J(\vec{w}, b) = \frac{1}{2m} \sum_{i=1}^{m} (f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)})^2 + \frac{\lambda}{2m} \sum_{j=1}^{n} w_j^2$$

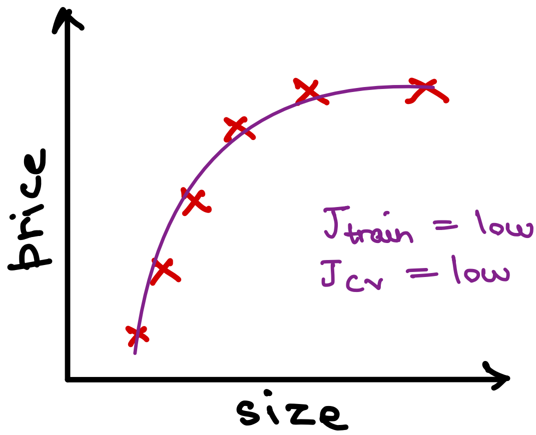


If $\lambda = 10{,}000$ (too high)
$f_{\vec{w},b}(\vec{x}) \approx b$,

because in case of λ being too high, the model makes the parameters very small - almost zero



Small λ (≈ 0) means regularization might not even occur which means it overfit (high variance)



Intermediate λ.
Regularization happens

If you're looking for perfect value for λ, the J_{cv} can be used

how?

Try multiple values for λ and see for which J_{cv} is the lowest.

1. Try $\lambda = 0 \rightarrow \min\limits_{\vec{w}, b} J(\vec{w}, b) \rightarrow w^{<1>}, b^{<1>} \rightarrow J_{cv}(w^{<1>}, b^{<1>})$

2. Try $\lambda = 0.01 \rightarrow \quad \rightarrow w^{<2>}, b^{<2>} \rightarrow J_{cv}(w^{<2>}, b^{<2>})$

3. Try $\lambda = 0.02 \rightarrow \quad \rightarrow w^{<3>}, b^{<3>} \rightarrow J_{cv}(w^{<3>}, b^{<3>})$

4. Try $\lambda = 0.04 \rightarrow \quad \rightarrow w^{<4>}, b^{<4>} \rightarrow J_{cv}(w^{<4>}, b^{<4>})$

⋮

12. Try $\lambda = 10 \rightarrow \quad \rightarrow w^{<12>}, b^{<12>} \rightarrow J_{cv}(w^{<12>}, b^{<12>})$

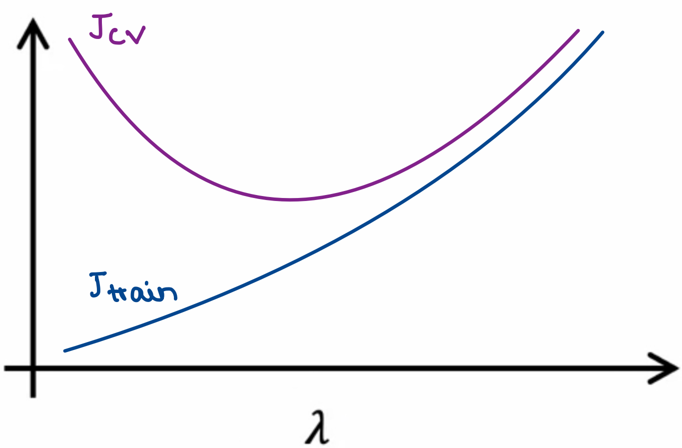
Check for which w, b is J_{cv} is the least

For example :-

If $w^{<5>}, b^{<5>}$ is trained on the training set and J_{cv} is the lowest, then check test error — $J_{test}(w^{<5>}, b^{<5>})$

To get even further intuition, let's see how λ affects training error and cross validation error.

$$J(\vec{w}, b) = \frac{1}{2m}\sum_{i=1}^{m}\left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right) - y^{(i)}\right)^2 + \frac{\lambda}{2m}\sum_{j=1}^{n} w_j^2$$



If $\lambda = 0$, J_{train} will be low as regularization doesn't even occur and keeps increasing. Whereas, J_{cv} is high, decreases and then keeps on increasing.