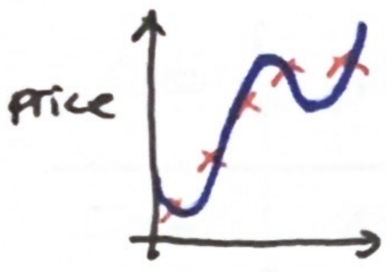
<Model Selection and Training/Validation/Test sets>



$f_{\vec{w},b}(\vec{x}) = w_1x_1 + w_2x^2 + w_3x^3 + w_4x^4 + b$

⇒ Once parameters $\vec{w}, b$ are fit to the training set, the training error $J_{train}(\vec{w},b)$ is likely lower than the actual generalization error.

⇒ $J_{train}(\vec{w},b)$: not good indicator of how well the model will likely do on new data

⇒ $J_{test}(\vec{w},b)$ is better estimate of how well the model will generalize to new data than $J_{train}(\vec{w},b)$

* Use test set to choose a model

d = degree of polynomial

fit parameters and get

d=1 : $f_{\vec{w},b}(\vec{x}) = w_1x_1 + b$ → $w^{<1>}, b^{<1>}$ —test→ $J_{test}(w^{<1>}, b^{<1>})$

d=2 : $f_{\vec{w},b}(\vec{x}) = w_1x_1 + w_2x^2 + b$ → $w^{<2>}, b^{<2>}$ —test→ $J_{test}(w^{<2>}, b^{<2>})$

d=3 : $f_{\vec{w},b}(\vec{x}) = w_1x_1 + w_2x^2 + w_3x^3 + b$

⋮

d=10 : $f_{\vec{w},b}(\vec{x}) = w_1x_1 + w_2x^2 + \cdots + w_{10}x^{10} + b$ → $w^{<10>}, b^{<10>}$ —test→ $J_{test}(w^{<10>}, b^{<10>})$

⇓

if find $J_{test}(w^{<5>}, b^{<5>})$ is the lowest

⇒ 이렇게 계산된 인덱의 $J_{test}()$ 중 d=5 일 때의 test error가 가장 작아서

$w_1x_1 + w_2x_2^2 + \cdots + w_5x_5^5 + b$ 을 모델로 선택했다고 하면,

이때 $J_{test}(w^{<5>}, b^{<5>})$ 가 model의 generalization을 가장 잘 반영한 것일까?

⇒ 사실 $J_{test}(w^{<5>}, b^{<5>})$ 는 generalization error의 optimistic estimate인 가능성이 높음
(= lower than the actual generalization error)

∵ An extra parameter d (degree of polynomial) was chosen using the test set.

↳ test set에 fit되었기 때문

∴ 주어진 데이터를 단순히 training/test로 나누는 것이 아닌 d를 fit하기 위한 set 역시 만들어야함.