ML Diagnostic:

Function ① 判断 Learning Algorithm 是否正确运行.

② 给出显著性改善 Learning Algorithm 的方法，并排除对其影响很小，甚至起负面作用的方法.

Learning Algorithm 常用 improve 方法 (Take linear regression as an example)

① Increasing features. → High variance (Overfit)

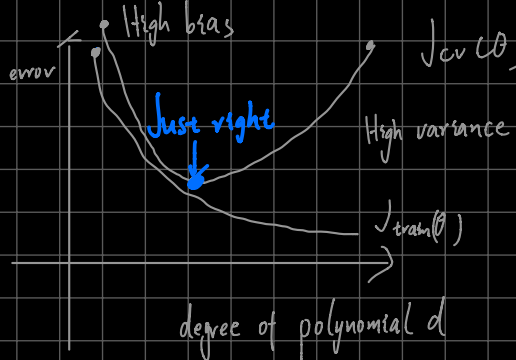
② Decreasing features → High variance (Overfit)

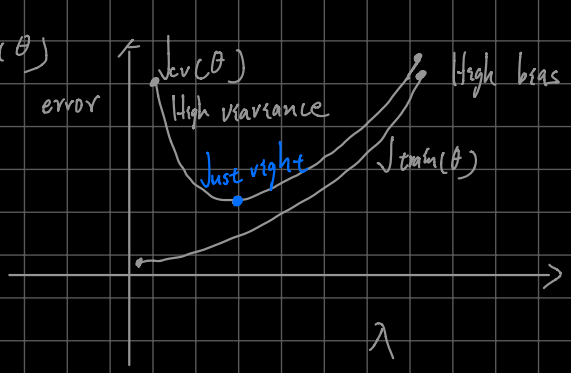
③ Adding polynomial features → High bias

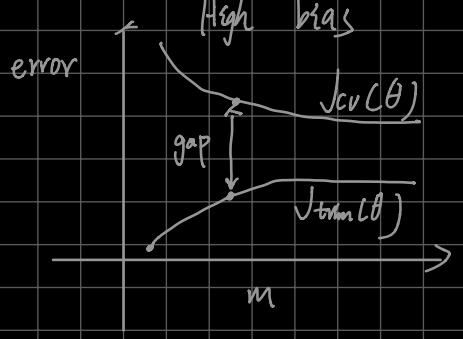
④ Increasing training samples → High variance (Overfit)

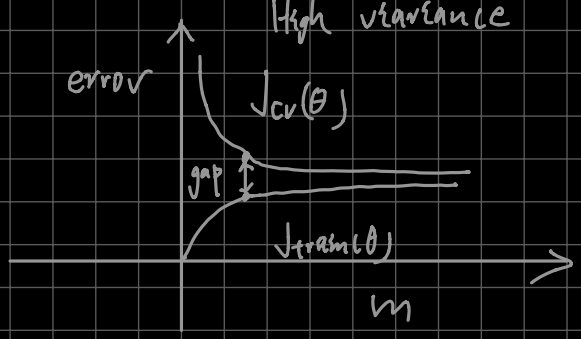
⑤ Increasing λ → High bias (Underfit)

⑥ Decreasing λ → High variance (Overfit)









Model Selection: (Take linear regression as an example)

select $d=1 \,||\, d=2 \,||\, \cdots \,||\, \underbrace{d=n}_{\text{degree of polynomial}}$

$\Rightarrow$ Use the CV data to choose the best model.

choose the model $(d=d_k)$ with lowest CV error.

△ Thus when in model selection we choose the best model with $d=d^*$, we can't use the test set to choose. That's why we split (often is 20%) out the CV data set to choose the best model.

actually we can view this procedure as determine an extra parameter d.

△

Error
- Training set errors
- CV set errors
- Generalization errors (Test set error)