Bias (viés):

* Que segue uma tendência ex: ‘pesquisa do departamento de trânsito pergunta se o motorista dirige de forma imprudente’, a resposta na maioria das vezes será não mesmo sendo imprudente.
* Erros de natureza sistemático.
* Pode ocorrer quando se tem poucos parâmetros causando underfitting (falta parâmetros) no treinamento do modelo.
* Ocorre com funções mais simples como o primeiro gráfico da esquerda para direita da Figura 01: uma regressão linear simples.

Variância

* Referente ao conjunto de dados de treinamento.
* Quando alta apresenta um treinamento com percentual de acerto alto e baixas taxas de erro, mas com baixo rendimento com o conjunto de teste (dados desconhecidos) tendo altas taxas de erro.
* Pode ocorrer quando se tem muitos parâmetros causando overfitting (excesso de parâmetros vindo de resíduos/ruídos) no treinamento do modelo.
* Ocorre com funções mais complexas como no terceiro gráfico da esquerda para a direita da Figura 01: uma regressão polinomial complexa.

Então se o modelo é muito simples e tem poucos parâmetros, tende a ter alto bias e baixa variância (primeiro gráfico da Figura 01) e se o modelo tende a ter muitos parâmetros e ser mais complexo tende a ter alta variância e baixo bias (terceiro gráfico da Figura 01).

O ideal então é achar um meio termo (fazendo regularização) onde teremos um balanceamento entre bias e variância (segundo gráfico da Figura 01): uma regressão polinomial de uma função mais simples, com menos parâmetros equalizando bias e variância.

A Figura 02 mostra a evolução do erro em função da complexidade do modelo (nro de variáveis da função):

* Inicialmente temos um modelo com bias baixo e alta variância, vemos que o erro de teste (ou generalização) acompanha o erro de treinamento, mas há ocorrência de underfitting.

Note que a complexidade do modelo está baixa fazendo que os erros de treino e de teste (generalização) sejam próximos.

* Temos na região central uma posição ideal para bias e variância, essa seria a condição viável para a complexidade do modelo.
* E agora a medida que a complexidade do modelo aumenta temos um baixo bias e uma alta variância, o erro de treinamento continua caindo como se esperava, mas o erro de teste ou generalização feito com dados desconhecidos do treinamento tem seu erro elevado, temos aqui um overfitting.

Mais adiante falaremos da regularização do modelo (função) para balancearmos bias e variância, até lá!



Figura 01

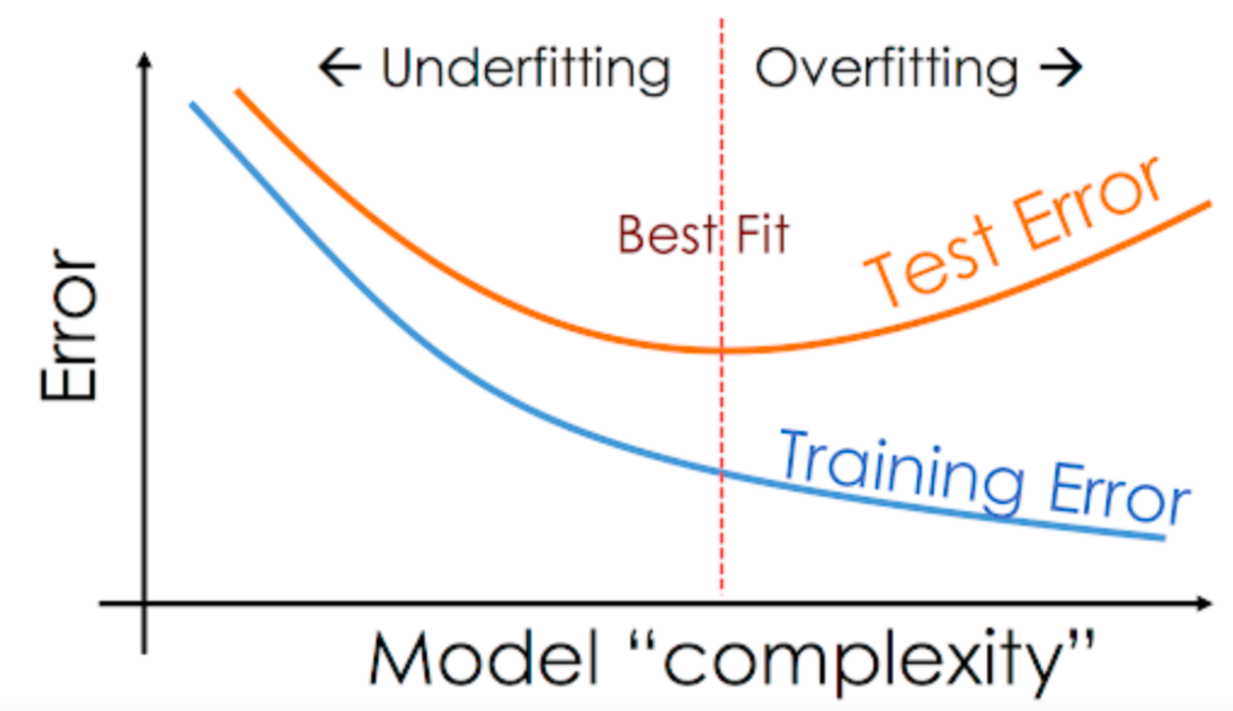


Figura 02