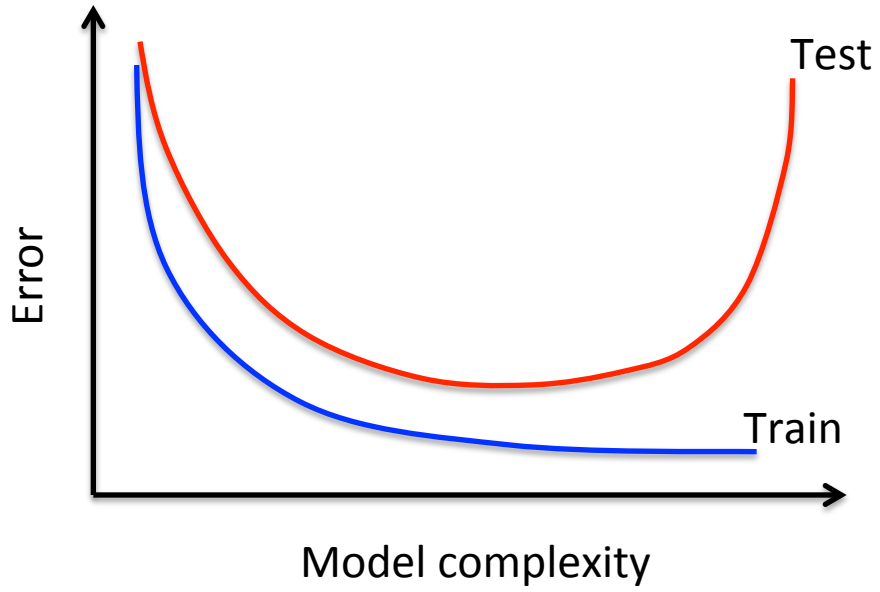
Review:
Model Selection

Training vs. Test errors

- Polynomial regression
  - Model complexity: Degree of polynomial
  - Is larger always better?



Model Selection Criterion

- How does once choose the 'best' polynomial degree using only the training set?
- Use a *model selection criterion* as a proxy for the test error:

-2 x Log-likehood + penalty term

Model Selection Criterion

- Akaike Information Criterion
  - AIC = -2 x Log-likehood + 2 x *K*
  - For least-squares regression:

K: degree of polynomial
n: training sample size

$$\text{AIC} = n \log\left(\frac{\text{RSS}}{n}\right) + 2K$$

- Bayesian Information Criterion (BIC)
  - BIC = -2 x Log-likehood + 2 x log(*K*)
  - For least-squares regression:

$$\text{BIC} = n \log\left(\frac{\text{RSS}}{n}\right) + \log(n)K$$

Note: The AIC and BIC definitions are slightly different from the text book, and correspond to the case where the residual error variance σ^2 is unknown.

Variable Selection

Exhaustive Search

- For each size ‘k’:
  - Enumerate all subsets of size ‘k’
  - Fit regression model for each subset
  - Pick subset with maximum R^2
- Use BIC to choose best size, and output optimal subset for that size

Enumerating Subsets

- Enumerate all subsets of predictors {0, 1, 2 , 3}
  - Subsets of size 1: {0}, {1}, {2}, {3}
  - Subsets of size 2: {0, 1}, {0, 2}, {0, 3},
    {1, 2}, {1, 3}, {2, 3}
  - Subsets of size 3: {0, 1, 2}, {0, 1, 3},
    {0, 2, 3}, {1, 2, 3}
  - Subsets of size 4: {0, 1, 2 , 3}

Enumerating Subsets

Best R^2 within each group

- Enumerate all subsets of predictors {0, 1, 2 , 3}
  - Subsets of size 1: {0}, {1}, {2}, {3} → Best 1-subset
  - Subsets of size 2: {0, 1}, {0, 2}, {0, 3}, {1, 2}, {1, 3}, {2, 3} → Best 2-subset
  - Subsets of size 3: {0, 1, 2}, {0, 1, 3}, {0, 2, 3}, {1, 2, 3} → Best 3-subset
  - Subsets of size 4: {0, 1, 2 , 3} → Best 4-subset

Enumerating Subsets

- Enumerate all subsets of predictors {0, 1, 2 , 3}
  - Subsets of size 1: {0}, {1}, {2}, {3} → Best 1-subset
  - Subsets of size 2: {0, 1}, {0, 2}, {0, 3}, {1, 2}, {1, 3}, {2, 3} → Best 2-subset
  - Subsets of size 3: {0, 1, 2}, {0, 1, 3}, {0, 2, 3}, {1, 2, 3} → Best 3-subset
  - Subsets of size 4: {0, 1, 2 , 3} → Best 4-subset

Choose subset with lowest BIC

Enumerating Subsets

- Generate all subsets of `set` of size `k`

```
subsets_k = itertools.combinations(set, k)
```

- Output is a list-like object

- Iterating through the generated subsets

```
for subset in subsets_k:
    ...
```

Putting it together

```
# Outer loop: iterate over sizes 1 …. d
for k in range(d):
    # Enumerate subsets of size ‘k’
    subsets_k = itertools.combinations(predictors, k)
```

Putting it together

```
# Outer loop: iterate over sizes 1 …. d
for k in range(d):
    # Enumerate subsets of size 'k'
    subsets_k = itertools.combinations(predictors, k)

    # Inner loop: iterate through subsets_k
    for subset in subsets_k :
        # Fit regression model using 'subset' and calculate R^2
        # Keep track of subset with highest R^2
        …
```

Putting it together

```
# Outer loop: iterate over sizes 1 …. d
for k in range(d):
    # Enumerate subsets of size 'k'
    subsets_k = itertools.combinations(predictors, k)

    # Inner loop: iterate through subsets_k
    for subset in subsets_k :
        # Fit regression model using 'subset' and calculate R^2
        # Keep track of subset with highest R^2
        …
```

Finds k-sized subset with best R^2

Putting it together

```
# Outer loop: iterate over sizes 1 …. d
for k in range(d):
    # Enumerate subsets of size 'k'
    subsets_k = itertools.combinations(predictors, k)

    # Inner loop: iterate through subsets_k
    for subset in subsets_k :
        # Fit regression model using 'subset' and calculate R^2
        # Keep track of subset with highest R^2
        …

    # Compute BIC of the subset you get from the inner loop
    # Compare with lowest BIC so far
```

Finds k-sized subset with best R^2