Principal Component Analysis

Background

Projecting $x^{(i)}$ onto unit vector v: $x_p^{(i)} = \underbrace{(x^{(i)T}v)}_{distance}\underbrace{v}_{direction}$

SVD decomposition: $X = USV^T$

Symmetric matrix $A = XX^T = (VSU^T)(USV^T) = VDV^T$

(S diagonal → SS = D still diagonal)

PCA ideas:

1. For each feature, computer the mean or zero center the training examples.
2. Find $k < d$ vectors in $\mathbb{R}^d$: $v^{(1)}, v^{(2)}, ..., v^{(k)}$ which are orthogonal unit (orthonormal) vectors.
3. Project the training examples.

PCA algorithm:

Input: the zero-centered data matrix X and $k \geq 1$

1. Compute the SVD of X: $[U, S, V] = svd(X)$
2. Choose the first k columns of V: $V_k = [v^{(1)}, ..., v^{(k)}]$
3. The PCA-feature matrix is $Z = XV_k$
   $\hat{X} = XV_kV_k^T$

PCA linearly projects examples into a lower dimensional space $N \times d$ into $N \times k$, $k < d$.

PCA maintains as much of the original variance (and minimizes least square reconstruction error).

Computing variance of the points $var = \frac{1}{N}\Sigma_{i=1}^{N}(x^{(i)T}v)^2 = \frac{1}{N}v^TX^TXv$

Let $A = X^TX = VDV^T$, we need to find $\arg\max_{v:||v||=1} v^TAv$

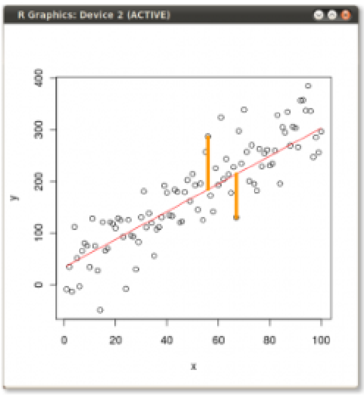
$v = Ve^{(1)} = \lambda_1$

PCA v.s. Ordinary Least Squares (OLS)

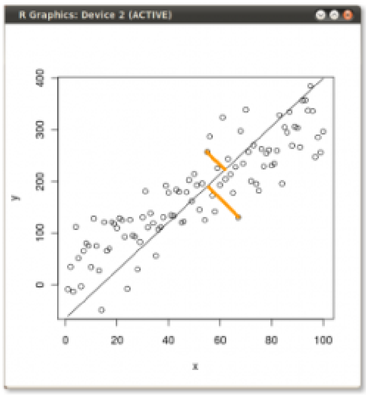
OLS — minimize perpendicular error

PCA — minimize orthogonal error

OLS


R Graphics: Device 2 (ACTIVE)
400
300
200
100
0
y
0
20
40
60
80
100
x

PCA


R Graphics: Device 2 (ACTIVE)
400
300
200
100
0
y
0
20
40
60
80
100
x