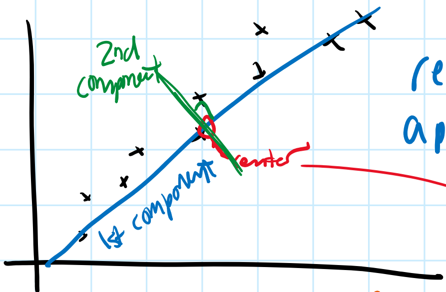
PCA

Tuesday, July 14, 2020 2:43 AM

Principal Component Analysis

* Reduce the rank of a matrix through approximation.



reality: rank 2
approx: rank 1

→ Get SVD of $\tilde{X} = \begin{bmatrix} — & x_1^T - \bar{x}^T & — \\ & \vdots & \\ — & x_m^T - \bar{x}^T & \end{bmatrix}$ (mean: $\bar{x}$)

$$\tilde{X} = U\Sigma V^* = \sum_{i=1}^{r} \sigma_i (u_i v_i^*)$$

Variance: $\frac{\sigma_i^2}{m-1}$ = average squared deviation from mean

PCA via Covariance Matrix

- Let X be an $n \times d$ matrix with n data points $\in \mathbb{R}^d$:

$$X = \begin{bmatrix} — & x_1^T & — \\ — & x_2^T & — \\ & \vdots & \\ — & x_n^T & \end{bmatrix}$$ ← one data point

- Centered data: $\tilde{X} = X - \underbrace{\frac{1}{n} \vec{1} \cdot \vec{1}^T X}_{\text{mean}}$

- Projection of data point x_1 on w:

$$\bar{x}_1 = \langle x_1, w \rangle w = (x_1 \cdot w) w = (x_1^T w) w$$

- $\underset{\text{maximize variance}}{\max}$ $\text{Var}\{\bar{x}_1, \bar{x}_2, \ldots \bar{x}_n\}) = \frac{1}{n} \sum_{i=1}^{n} \text{Var}(\bar{x}_i)$

$$= \max_{\vec{w}} \frac{1}{n} \sum_{i=1}^{n} \underbrace{(\langle \tilde{x}_i, w \rangle)^2}_{\text{var}(\bar{x}_i)}$$ ← mean squared error: avoid cancellations

Maximize variance $= \max_{\vec{w}} \frac{1}{n} \sum_{i=1}^{n} \underbrace{(\langle \tilde{x}_i, w \rangle)^2}_{var(\tilde{x}_i)}$ ← mean squared error: cancellations

$$= \max_{\vec{w}} \frac{1}{n} \sum_{i=1}^{n} (\tilde{x}_i^T w^2) = \max_{\vec{w}} \frac{1}{n} \|\tilde{X} w\|^2$$

$$= \max_{\vec{w}} \frac{1}{n} \underbrace{(\tilde{X} w)^T (\tilde{X} w)}_{= w^T \tilde{X}^T \tilde{X} w}$$

w = eigenvector corresponding to λ_i; $\lambda_{max}(\tilde{X}^T \tilde{X}) = \lambda_1$ — positive eigenvalues, orthogonally diagonalizable

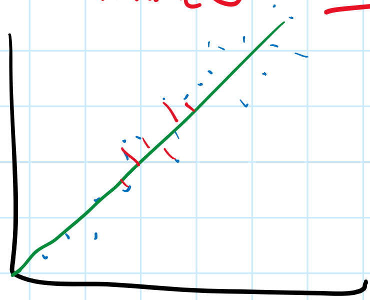
- Covariance matrix: $S = \frac{1}{n-1} \tilde{X}^T \tilde{X}$ (average value of XX^*)

→ How to compute PCA as shown above: diagonalize the covariance matrix.

↳ Subtract the component of X_n along first component, then apply $\tilde{X}^T \tilde{X}$ again to get 2nd component

PCA via SVD

→ idea: minimize the error:



difference from least squares: project orthogonally, not vertically

$$\min \left(\sum_{i=1}^{n} \|\tilde{x}_i - \bar{x}_i\|^2 \right)$$

$$= \min \sum_{i=1}^{n} \| \tilde{x}_i - (x_i^T w) w \|^2$$

$$= \min \sum_{i=1}^{n} (\tilde{x}_i - (\tilde{x}_i^T w) w)^T (\tilde{x}_i - (\tilde{x}_i^T w) w) \quad \rightleftarrows \quad \|v\|^2 = v^T v$$

$$\min \sum_{i=1} (x_i - (x_i^T w) w)^T (x_i - (x_i^T w) w)$$

$$= \min \sum_{i=1}^{n} \left(\tilde{x}_i^T - (x_i^T w) w^T \right) \left(x_i - (x_i^T w) w \right)$$

$$= \underbrace{\min \sum_{i=1}^{n} \Big(\underbrace{\|\tilde{x}_i\|^2}_{\text{constant}} - (\tilde{x}_i^T w)^2}_{n \cdot \mathrm{Var}(x_1) \text{ from prev. section}} \Big)$$

Minimizing error = maximizing variance

$$\min \sum_{i=1}^{n} -n \cdot \mathrm{Var}(x_1)$$

$$= \max \left(\mathrm{Var}(\tilde{X}) \right)$$

Computing PCA

1) Compute $A = \frac{1}{m} \vec{1}\vec{1}^T X$.

2) Compute covariance matrix $S = \frac{1}{m} A^T A$

3) Diagonalize $S = P \Lambda P^T$. The cols of P are principal components.
The weights are the singular values.