Applied Machine Learning

SVD and NIPALS

SVD and NIPALS

- Singular Value Decomposition (SVD)
- Non-linear Iterative Partial Squares (NIPALS)

Principal Component Analysis

- Original dataset: $\{\mathbf{x}\}$
  - d features
  - $U^\top \text{Covmat}\{\mathbf{x}\}U = U^\top \Sigma U = \Lambda$
- Choose s features
  - small ratio in $\dfrac{\sum_{j=s+1}^{j=d} \lambda_j}{\sum_{j=1}^{j=d} \lambda_j}$
  - plot relative error vs s or λ_i vs i, select s most significant

- Low dimensional representation
  - New dataset: $\{\hat{\mathbf{x}}\}$
  - $\hat{\mathbf{x}}_i \quad = \quad \sum_{j=1}^{j=s} \left[\mathbf{u}_j^\top (\mathbf{x}_i - \text{mean}(\{x\}))\mathbf{u}_j\right] + mean(\{\mathbf{x}\})$
  - lower rank
- Singular Value Decomposition
- Non-linear Iterative Partial Squares

Singular Value Decomposition

- Decompose $X_{m\times p}$ as $U_{m\times m}S_{m\times p}V^{\top}_{p\times p}$
- $X = USV^{\top}$
  - U: orthonormal $UU^{\top} = I$
  - S: diagonal matrix
    - Singular values
  - V: orthonormal $VV^{\top} = I$
- Elements in S in descending order
- U and $V^{\top}$ sorted accordingly
- Use numerical libraries

SVD and Covariance Matrix

- $X^\top X$
  - Symmetric:: $(X^\top X)^\top = (X)^\top (X^\top)^\top = X^\top X$
  - applying SVD:
    $$\begin{aligned} X^\top X &= (USV^\top)^\top (USV^\top) \\ &= (V^\top)^\top S^\top U^\top USV^\top \\ &= VS^\top SV^\top \\ &= VS^2V^\top \end{aligned}$$
- Dataset with zero mean:
  - $$X = \begin{bmatrix} x_1^\top \\ x_2^\top \\ \vdots \\ x_N^\top \end{bmatrix} \quad \begin{aligned} \text{Covmat}\{X\} &= \frac{\sum_i (x_i - mean(\{X\}))(x_i - mean(\{X\}))^\top}{N} \\ \Sigma &= \frac{\sum_i x_i x_i^\top}{N} \\ &= \frac{1}{N} X^\top X \end{aligned}$$

- $$\begin{aligned} \Sigma &= \frac{1}{N} X^\top X \\ &= \frac{1}{N} VS^2V^\top \end{aligned}$$
- $$\begin{aligned} \Sigma V &= \frac{1}{N} VS^2V^\top V \\ \Sigma V &= V\frac{S^2}{N} \end{aligned}$$
  - V: eigenvectors of Σ
  - $\Lambda = \dfrac{S^2}{N}$: eigenvectors of Σ
    - matrix of variances of X

NIPALS

- Dataset with zero mean:
  - $$X = \begin{bmatrix} x_1^\top \\ x_2^\top \\ \vdots \\ x_N^\top \end{bmatrix}$$
- First principal component. Find:
  - $\mathbf{u}$ and $\mathbf{w}$
  - $\|\mathbf{u}\| = 1$
  - $\mathbf{w}\mathbf{u}^\top$ the closest to X
- $\|\mathbf{w}\|$
  - largest singular value
- $\mathbf{u}^\top$
  - row of $V^\top$ corresponding to largest singular value
- $\dfrac{\mathbf{w}}{\|\mathbf{w}\|}$
  - column of U corresponds to largest singular value

NIPALS

- $\mathbf{w}\mathbf{u}^\top$ the closest to X
  - Frobenius norm for matrix A: $\|A\|_F^2 = \sum_{i,j} a_{i,j}^2$
- Cost function to minimize: $C(\mathbf{x}, \mathbf{u}) = \|X - \mathbf{w}\mathbf{u}^\top\|_F^2 = \sum_{i,j} (x_{i,j} - w_i u_j)^2$ $\quad\quad \|\mathbf{u}\| = 1$
  - $\frac{\partial C}{\partial w_i} \quad = \quad 2\sum_j (x_{i,j} - w_i u_j) u_j \qquad \frac{\partial C}{\partial u_j} \quad = \quad 2\sum_j (x_{i,j} - w_i u_j) w_i$
  - $\nabla_w C \quad = \quad (X - \mathbf{w}\mathbf{u}^\top)\mathbf{u} \quad = 0 \qquad \nabla_u C \quad = \quad (X^\top - \mathbf{u}\mathbf{w}^\top)\mathbf{w} \quad = 0$

  $\hat{\mathbf{w}} \quad = \quad \frac{X\mathbf{u}}{\mathbf{u}^\top\mathbf{u}} \qquad \hat{\mathbf{u}} \quad = \quad \frac{X^\top\mathbf{w}}{\mathbf{w}^\top\mathbf{w}}$

  normalization: $\|\mathbf{u}\| = 1$:

  - $\mathbf{w} \quad = \quad \left(\sqrt{\hat{\mathbf{u}}^\top\hat{\mathbf{u}}}\right)\hat{\mathbf{w}} \qquad \mathbf{u} \quad = \quad \frac{\hat{\mathbf{u}}}{\sqrt{\hat{\mathbf{u}}^\top\hat{\mathbf{u}}}}$

NIPALS: Algorithm

1. $\mathbf{u}^{(0)}$ initialized at random and normalize $\|\mathbf{u}^{(0)}\| = 1$
2. For ($n = 1; \|\mathbf{u}^{(n)} - \mathbf{u}^{(n-1)}\| > \epsilon; n++$)
   1. $\hat{\mathbf{w}} = \dfrac{X\mathbf{u}^{(n-1)}}{\mathbf{u}^{(n-1)\top}\mathbf{u}^{(n-1)}}$
   2. $\hat{\mathbf{u}} = \dfrac{X^\top \hat{\mathbf{w}}}{\hat{\mathbf{w}}^\top \hat{\mathbf{w}}}$
   3. $\mathbf{w}^{(n)} = \left(\sqrt{\hat{\mathbf{u}}^\top \hat{\mathbf{u}}}\right)\hat{\mathbf{w}} \qquad \mathbf{u}^{(n)} = \dfrac{\hat{\mathbf{u}}}{\sqrt{\hat{\mathbf{u}}^\top \hat{\mathbf{u}}}}$

- 1st principal component:
  - $\|\mathbf{w}\|$: singular value
  - $\mathbf{u}^\top$: row of $V^\top$
  - $\dfrac{\mathbf{w}}{\|\mathbf{w}\|}$: column of U
- 2nd principal component:
  - Apply NIPALS to $X^{(1)} = X - \mathbf{w}\mathbf{u}^\top$
- Low rank representation: $\hat{\mathbf{x}}_i^\top = \sum_{j=1}^{s} w_{i,j}\mathbf{u}_j^\top$

NIPALS

- Missing values
  - Formulate updates as sums instead of Matrix operations, ignore missing (i, j)
- Smooths Gaussian noise
- May smooth noise that comes from
  - counts
  - missing entries
- Available in numerical libraries

SVD and NIPALS

- Singular Value Decomposition (SVD)
- Non-linear Iterative Partial Squares (NIPALS)

Applied Machine Learning

SVD and NIPALS