מבוא ללמידה חישובית- תרגול אחרון

סיכום הקורס

אלגוריתמים

Unsupervised

Clustering:

- K-means
- Gaussian mixture model (EM)
- Spectral Clustering

Dim. Reduction

- (kernel) PCA

Supervised

Binary Classification:

- Naïve Bayes
- Hyperplane algorithms
  - (Hard) SVM
  - Soft Margin SVM
  - Kernel SVM
    - לינארי
    - פולינומיאלי
    - גאוסיאני
    - ...
  - (Margin +) Perceptron
  - Winnow
- Logistic Regression (עם עיגול)
- k-NN (+ kernel)
- Decision Trees
- Boosting
  - Random Forest
  - Stacking, Bagging
  - AdaBoost

Multiclass Classification:

- אסטרטגיות להמרה של אלגוריתם שלומד בינארית לאלגוריתם שלומד קטגוריאלית:
  - One VS. All
  - One VS. One
- k-NN
- Naïve Bayes

Regression:

- Linear Regression
  - Lasso
  - Ridge
- Logistic Regression
- SVR (SVM for regression)

כלים

אופטימיזציה

- EM
- Gradient Descent
- Coordinate Descent
- Linear Programming
- Quadratic Programming

סטטיסטיקה

- Maximum Likelihood
- MAP- Maximum A-posteriori estrimator

אחרים

- PAC
- SVD

PCA

$$\bar{x}_1, \dots, \bar{x}_n \in \mathbb{R}^d$$

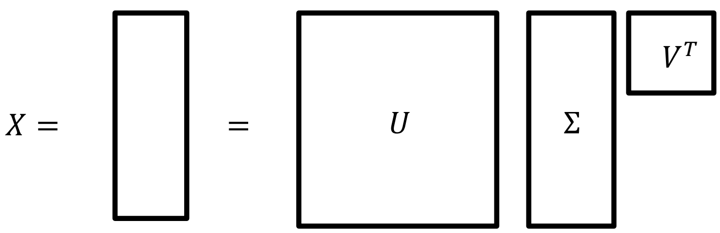
$$X = \underbrace{\left.\begin{pmatrix} | & \cdots & | \\ \bar{x}_1 & \dots & \bar{x}_n \\ | & \cdots & | \end{pmatrix}\right\} d}_{n}$$

מניחים: $\sum_{i=1}^{n} \bar{x}_i = \bar{0}$

PCA: מחשב וקטורים עצמיים של XX^T.

סיבוכיות: $O(d^2 \cdot n + d^3)$.

מה קורה אם $n \ll d$?



נעשה פירוק לפי SVD:

$$X = U\Sigma V^T$$

ונרצה את $\bar{u}_1, \dots, \bar{u}_n \in \mathbb{R}^d$

$$XX^T = (U\Sigma V^T)(V\Sigma^T U^T) = U \cdot diag(\sigma_1^2, \dots, \sigma_n^2, 0, \dots, 0) \cdot U^T$$
$$X^T X = V \cdot diag(\sigma_1^2, \dots, \sigma_n^2) \cdot V^T$$

נניח שפירקנו ספקטרלית את $X^T X$:

- קיבלנו את V (כלומר $\bar{v}_1, \dots, \bar{v}_n \in \mathbb{R}^n$)
- ואת $\sigma_1^2, \dots, \sigma_n^2$

<u>ניזכר ש</u>:

$$X\bar{v}_i = (U\Sigma V^T)\bar{v}_i = U\Sigma\bar{e}_i = U \cdot (\sigma_i \cdot \bar{e}_i) = \sigma_i \cdot U\bar{e}_i = \sigma_i \cdot \bar{u}_i$$

לכן:

$$X\bar{v}_i = \sigma_i \cdot \bar{u}_i$$

אז אפשר לחשב את $\bar{u}_1, \dots, \bar{u}_n$ כך:

$$\bar{u}_i = \frac{1}{\sigma_i} \cdot X\bar{v}_i$$

<u>סיבוכיות</u>:

חישוב $X^T X$: $O(n^2 \cdot d)$

פירוק ספקטרלי: $O(n^3)$
חישוב $\bar{u}$ מתוך $\bar{v}$: $O(n^2 \cdot d)$
סך הכול: $O(n^2 \cdot d + n^3 + n^2 \cdot d) = O(n^2 \cdot d + n^3)$

<u>שאלה ממבחן</u>

א. נתונה בעיית רגרסיה, שמקבילה לבעיית האופטימיזציה:

$$\min_{a,b} \sum_{i=1}^{n} |y - ax_i - b|$$

כאשר $x_1, \dots, x_n, y \in \mathbb{R}$ וגם $a, b \in \mathbb{R}$. לאיזה מודל הסתברותי הבעייה מתאימה?

<u>תשובה</u>: המודל-

$$y_i = ax_i + b + \varepsilon_i$$
$$\varepsilon_i \sim Laplace(0,1)$$
$$\Pr(\varepsilon_i = x) = \frac{1}{2} e^{-|x|}$$

ואז Maximum Likelihood של המודל שקול לבעיית המינימיזציה.

הסבר:

$$logL(a,b;\{x_i,y_i\}) = \sum_{i=1}^{n} \log\left(\Pr_{a,b}(y_i|x_i)\right) = \sum_{i=1}^{n} \log\left(\Pr_{a,b}(y_i - ax_i - b)\right) =$$

$= \varepsilon_i$ ולכן מתפלג לפלס

$$\sum_{i=1}^{n} \log\frac{1}{2} - |y_i - ax_i - b|$$

ב. איך פותרים? עם LP?
נמציא משתנים חדשים:

$$\gamma_i = y - ax_i - b$$

ואז:

$$\min_{\mathrm{a,b,\gamma}} \sum_{i=1}^{n} \gamma_i$$
$$s.t.$$
$$\gamma_i \geq y - ax_i - b$$
$$\gamma_i \geq -(y - ax_i - b)$$