QP & cone program duality
Support vector machines

10-725 Optimization
Geoff Gordon
Ryan Tibshirani
Review

• Quadratic programs

• Cone programs
  ‣ SOCP, SDP
  ‣ QP ⊆ SOCP ⊆ SDP
  ‣ SOC, S+ are self-dual

• Poly-time algos (but not strongly poly-time, yet)

• Examples: group lasso, Huber regression, matrix completion
Matrix completion

- Observe $A_{ij}$ for $ij \in E$, write $O_{ij} = \begin{cases} 1 & \text{if } \mathbf{0} \\ 0 \text{ otherwise} \end{cases}$

- $\min_X ||(X-A)\circ P||_F^2 + \lambda ||X||_*$

\[
X = U\Sigma V^T \\
M \succeq 0 \iff \text{tr}(B^T M) \geq 0 \text{ for all } B \succeq 0 \text{ take } B = \begin{pmatrix} uu^T & -uv^T \\ -v^TU & vv^T \end{pmatrix} \\
M = \begin{pmatrix} P & X \\ X^T & Q \end{pmatrix} \succeq 0
\]

\[
\lambda \left(\text{tr}(P) + \text{tr}(Q)\right) \geq 2 \text{tr}\left(U^T X V\right) = 2 \|X\|_F
\]

\[
P = U\Sigma U^T, \quad Q = V\Sigma V^T \\
\text{tr}(P) = \text{tr}(U^T U) + \text{tr}(U^T V) = 2\|X\|_F
\]
Max-variance unfolding
aka semidefinite embedding

- **Goal:** given $x_1, \ldots, x_T \in \mathbb{R}^n$
  - find $y_1, \ldots, y_T \in \mathbb{R}^k$ (\(k \ll n\))
  - \(\|y_i - y_j\| \approx \|x_i - x_j\| \quad \forall i, j \in E\)

- If $x_i$ were near a $k$-dim subspace of $\mathbb{R}^n$, PCA!

- Instead, two steps:
  - first look for $z_1, \ldots, z_T \in \mathbb{R}^n$ with
    - \(\|z_i - z_j\| = \|x_i - x_j\| \quad \forall i, j \in E\)
    - and var($z$) as big as possible
  - then use PCA to get $y_i$ from $z_i$
\begin{itemize}
\item \textbf{MVU/SDE}
\end{itemize}

\[ \text{max}_z \, \text{tr}(\text{cov}(z)) \text{ s.t. } ||z_i - z_j|| = ||x_i - x_j|| \quad \forall i, j \in E \]
Result

- Embed 400 images of a teapot into 2d

Euclidean distance from query to A is smaller; after MVU, distance to B is smaller.
Duality for QPs and Cone Ps

- Combined QP/CP:
  - \( \min c^T x + x^T H x / 2 \) s.t. \( Ax + b \in K \quad x \in L \)
  - cones \( K, L \) implement any/all of equality, inequality, generalized inequality
  - assume \( K, L \) proper (closed, convex, solid, pointed)
Primal-dual pair

- **Primal:**
  - \( \min c^T x + x^T H x / 2 \) \( \text{ s.t. } Ax + b \in K \) \( x \in L \)

- **Dual:**
  - \( \max -z^T H z / 2 - b^T y \) \( \text{ s.t. } Hz + c - A^T y \in L^* \) \( y \in K^* \)
KKT conditions

- primal pair
  - \( \min c^T x + \frac{x^T H x}{2} \) s.t. \( A x + b \in K \), \( x \in L \)
  - \( \max -b^T y - \frac{z^T H z}{2} \) s.t. \( H z + c - A^T y \in L^* \), \( y \in K^* \)
KKT conditions

- primal: $Ax+b \in K \quad x \in L$
- dual: $Hz + c - A^Ty \in L^* \quad y \in K^*$
- quadratic: $Hx = Hz$
- comp. slack: $y^T(Ax+b) = 0 \quad x^T(Hz+c-A^Ty) = 0$
Support vector machines
(separable case)
Maximizing margin

- margin $M = y_i (x_i \cdot \overline{w} - \overline{b})$
- $\max M \text{ s.t. } M \leq y_i (x_i \cdot \overline{w} - \overline{b})$
For example
\[ \text{min} \ ||v||^2/2 \quad \text{s.t.} \quad y_i (x_i^T v - d) \geq 1 \quad \forall i \]
SVM duality

- $\min ||v||^2/2 - \sum s_i \quad \text{s.t.} \quad y_i (x_i^T v - d) \geq 1 - s_i \quad \forall i$

- $\min v^T v/2 + l^T s \quad \text{s.t.} \quad Av - yd + s - l \geq 0$
Interpreting the dual

\[ \max l^T \alpha - \alpha^T K \alpha / 2 \quad \text{s.t.} \quad y^T \alpha = 0 \quad 0 \leq \alpha \leq 1 \]

\( \alpha^\prime \):
 \( \alpha > 0 \):
 \( \alpha < 1 \):
 \( y^T \alpha = 0 \):
From dual to primal

- \max_\alpha 1^T\alpha - \alpha^TK\alpha/2 \quad \text{s.t.} \quad y^T\alpha = 0, \quad 0 \leq \alpha \leq 1
A suboptimal support set
SVM duality: the applet
Why is the dual useful?

aka the kernel trick

\[
\text{max } 1^T \alpha - \alpha^T A A^T \alpha / 2 \quad \text{s.t. } y^T \alpha = 0 \quad 0 \leq \alpha \leq 1
\]

• SVM: n examples, m features
  ‣ primal:
  ‣ dual:
Ball center
aka Chebyshev center

- $X = \{ x \mid Ax + b \geq 0 \}$

- Ball center:
  - if $\|a_i\| = 1$
  - in general:
Analytic center

- Let $s = Ax + b$
- Analytic center:
Bad conditioning? No problem.
Newton for analytic center

- Lagrangian \( L(x,s,y) = -\sum \ln s_i + y^T(s-Ax-b) \)
Adding an objective

- Analytic center was for \{ x | Ax + b = s \geq 0 \}
- Now: \min c^T x \text{ s.t. } Ax + b = s \geq 0
- Same trick:
  - \min t c^T x - \sum \ln s_i \text{ s.t. } Ax + b = s \geq 0
    - parameter \ t \geq 0
  - central path =
    - \ t \to 0: \quad t \to \infty:
    - L(x,s,y) =
Newton for central path

- \( L(x,s,y) = t \ c^T x - \sum \ln s_i + y^T(s-Ax-b) \)