Machine Learning of Andrew Ng
pdf by CarolusRex

SVM-2

- SVM (support vector machine)
  - SMO
  - 非线性分类器
  - SMO derivation(concrete)
    - iterations choice

SVM (support vector machine)

本文着眼于SMO原理和非线性分类器。

SMO

网页。
网页。
platt论文 Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines。

非线性分类器

一个思路是将低维数据投射到高维数据，在高维空间中寻找超平面。

则代价函数由$\overrightarrow{x_i} \cdot \overrightarrow{x_j}$变为$\langle \phi(x^{(i)}), \phi(x^{(j)}) \rangle$，那么对于$k$维，其时间复杂度是$O(k^2)$（需要转化为高维空间），难以接受。一个可行$trick$是使用$kernel\ function$，其高维点积为低维点积转换后相乘，$\overrightarrow{h(x_i)} \cdot \overrightarrow{h(x_j)}$，则时间复杂度为$O(k)$。

常用$kernel$:

Kernel	expression
Linear	$K(x,y) = x^T y + c$
Polynomial	$K(x,y) = (ax^T y + c)^d, (a, c \geqslant 0)$
Radial Basis	$K(x,y) = exp(-\gamma \|x - y\|^2), (\gamma \geqslant 0)$
Gaussiaan	$K(x,y) = exp(-\frac{\|x-y\|^2}{2\sigma^2})$

Valid Kernel: 半正定对称矩阵。证明。

SMO derivation(concrete)

We have $\alpha_i y_i = 0$, so we have to change α_i, α_j simultaneously. Assume we choose α_1, α_2, then $\alpha_1 y_1 + \alpha_2 y_2 = -\sum_{i=3} \alpha_i y_i = \zeta$.

target:
$min\ L = \frac{1}{2}\sum_{i,j} \alpha_i \alpha_j y_i y_j < \overrightarrow{x_i}, \overrightarrow{x_j} > - \sum \alpha_i$

$$
\begin{aligned}
L =& \frac{1}{2}\alpha_1^2 K_{11} + \frac{1}{2}\alpha_2^2 K_{22} + \alpha_1 \alpha_2 y_1 y_2 K_{12} + \alpha_1 y_1 \sum_{i=3} \alpha_i y_i K_{i1} + \alpha_2 y_2 \sum_{i=3} \alpha_i y_i K_{i2} - (\alpha_1 + \alpha_2) + const \\
\alpha_1 =& y_1 \zeta - y_1 y_2 \alpha_2 \\
\frac{\partial \alpha_1}{\partial \alpha_2} =& - y_1 y_2 \\
\frac{\partial L}{\partial \alpha_2} =& \alpha_1 K_{11} \frac{\partial \alpha_1}{\partial \alpha_2} + \alpha_2 K_{22} + y_1 y_2 K_{12} \frac{\partial \alpha_1 \alpha_2}{\partial \alpha_2} + \frac{\partial \alpha_1}{\partial \alpha_2} y_1 \sum_{i=3} \alpha_i y_i K_{i1} + y_2 \sum_{i=3} \alpha_i y_i K_{i2} - 1 - \frac{\partial \alpha_1}{\partial \alpha_2} \\
=& - y_1 y_2 \alpha_1 K_{11} + \alpha_2 K_{22} + y_1 y_2 K_{12}(\alpha_1 - y_1 y_2 \alpha_2) - y_2 \sum_{i=3} \alpha_i y_i K_{i1} + y_2 \sum_{i=3} \alpha_i y_i K_{i2} + y_1 y_2 - 1 \\
=& (K_{11} + K_{22} - 2K_{12})\alpha_2 - y_2 K_{11} \zeta + y_2 K_{12} \zeta + y_1 y_2 - 1 - y_2 \sum_{i=3} \alpha_i y_i (K_{i1} - K_{i2})
\end{aligned}
$$

let $\frac{\partial L}{\partial \alpha_2} = 0$, then $(K_{11} + K_{22} - 2K_{12})\alpha_2 = y_2((K_{11} - K_{12})\zeta + y_2 - y_1 + \sum_{i=3} \alpha_i y_i (K_{i1} - K_{i2}))$

$$
\begin{aligned}
(K_{11} + K_{22} - 2K_{12})\alpha_2 =& y_2((K_{11} - K_{12})\zeta + y_2 - y_1 + \sum_{i=3} \alpha_i y_i (K_{i1} - K_{i2})) \\
(K_{11} + K_{22} - 2K_{12})\alpha_2 =& y_2(\sum \alpha_i y_i K_{i1} - \sum \alpha_i y_i K_{i2} + y_2 - y_1 + \alpha_2 y_2 (K_{11} + K_{22} - 2K_{12})) \\
(K_{11} + K_{22} - 2K_{12})\alpha_2^* =& (K_{11} + K_{22} - 2K_{12})\alpha_2 + y_2((\sum \alpha_i y_i K_{i1} - y_1) - (\sum \alpha_i y_i K_{i2} - y_2))
\end{aligned}
$$

let $E_i = \sum_j \alpha_i y_i K_{ij} + b - y_i, \eta = K_{11} + K_{22} - 2K_{12}$
$\alpha_2^* = \alpha_2 + \frac{y_2(E_1 - E_2)}{\eta}$

α_2^* also needs to satisfy $[L, H]$

$$
\alpha_2^{new} = \begin{cases} H, & (H < \alpha_2^*) \\ \alpha_2^*, & (L \leqslant \alpha_2^* \leqslant H) \\ L, & (\alpha_2^* < L) \end{cases}
$$

$\alpha_1^{new} = y_1(\eta - y_2 \alpha_2^{new}) = y_1(y_1 \alpha_1 + y_2 \alpha_2 - y_2 \alpha_2^{new}) = \alpha_1 + y_1 y_2 (\alpha_2 - \alpha_2^{new})$

iterations choice

$u = \sum y_j \alpha_j K(\overrightarrow{x_j}, \overrightarrow{x}) - b$
KKT condition of QP problem:

$$
\begin{aligned}
\alpha_i = 0 &\Leftrightarrow y_i u_i \geqslant 1 \\
0 < \alpha_i < C &\Leftrightarrow y_i u_i = 1 \\
\alpha_i = C &\Leftrightarrow y_i u_i \leqslant 1
\end{aligned}
$$

first choice the point which violates **KKT**, then choice the point of max $\|E_2 - E_1\|$.