Machine Learning of Andrew Ng
pdf by CarolusRex

SVM (support vector machine)

features

间隔最大的**线性分类器**
（感知机为可行NN，无间隔要求），
通过使用核技巧，可以进阶为非线性分类器。

本文着眼于间隔最大的线性分类器。

definition (hard margin)

hard margin满足严格分类。
对点集$\{(\overrightarrow{x_i}, y_i)\}$，令$\overrightarrow{w} \cdot \overrightarrow{x} + b = 0$为超平面，则几何间隔为$\gamma_i = y_i(\frac{\overrightarrow{w}}{||\overrightarrow{w}||} \cdot \overrightarrow{x_i} + \frac{b}{||\overrightarrow{w}||})$。

geometric margin derivation

记$\overrightarrow{x_0}$为$\overrightarrow{x}$在超平面上的投影，则有$\overrightarrow{x} = \overrightarrow{x_0} + \gamma\frac{\overrightarrow{w}}{||\overrightarrow{w}||}$，且满足$\overrightarrow{w} \cdot \overrightarrow{x_0} + b = 0$，
代入得，$\overrightarrow{w}(\overrightarrow{x} - y\frac{\overrightarrow{w}}{||\overrightarrow{w}||}) + b = 0$，解出$y = \frac{\overrightarrow{w} \cdot \overrightarrow{x} + b}{||\overrightarrow{w}||}$，
$\tilde{\gamma} = y\gamma = \frac{\hat{\gamma}}{||\overrightarrow{w}||}$

对一组固定的$pair(\overrightarrow{w},\ b)$，我们可以得到其对应的一组$\{\gamma_i\}$，则其间隔为$\gamma = min_i\ \gamma_i$。

target

target: $max_{\overrightarrow{w},b}\ \gamma$
由*def*可知，$\gamma_i \geqslant \gamma$，即$y_i(\frac{\overrightarrow{w}}{||\overrightarrow{w}||} \cdot \overrightarrow{x_i} + \frac{b}{||\overrightarrow{w}||}) \geqslant \gamma$，进一步的，有$y_i(\frac{\overrightarrow{w}}{||\overrightarrow{w}||\gamma} \cdot \overrightarrow{x_i} + \frac{b}{||\overrightarrow{w}||\gamma}) \geqslant 1$。

令$\overrightarrow{W} = \frac{\overrightarrow{w}}{||\overrightarrow{w}||\gamma}$，$B = \frac{b}{||\overrightarrow{w}||\gamma}$，则$y_i(\overrightarrow{W} \cdot \overrightarrow{x_i} + B) \geqslant 1$。
$||\overrightarrow{W}|| = \frac{||\overrightarrow{w}||}{||\overrightarrow{w}||\gamma} = \frac{1}{\gamma}$，则$max\ \gamma \Leftrightarrow max\ \frac{1}{||\overrightarrow{W}||} \Leftrightarrow min\ \frac{1}{2}||\overrightarrow{W}||^2$

在下文中，用$\overrightarrow{w}$代替$\overrightarrow{W}$，用b代替B，需要注意。

target: $min_{\overrightarrow{w},b}\ \frac{1}{2}||\overrightarrow{w}||^2$，$s.t.\ y_i(\overrightarrow{w} \cdot \overrightarrow{x_i} + b) \geqslant 1$
这是一个含有不等式约束的凸二次规划问题，考虑使用拉格朗日乘子进和dual problem。
构造无约束拉格朗日目标函数，$L(\overrightarrow{w}, b, \overrightarrow{\alpha}) = \frac{1}{2}\overrightarrow{w}^2 - \sum \alpha_i(y_i(\overrightarrow{w} \cdot \overrightarrow{x_i} + b) - 1)$，
记$\theta(\overrightarrow{w}, b) = max_{\alpha_i \geqslant 0}\ L(\overrightarrow{w}, b, \overrightarrow{\alpha}) = \begin{cases} \frac{1}{2}\overrightarrow{w}^2, (\forall\ i,\ y_i(\overrightarrow{w} \cdot \overrightarrow{x_i} + b) \geqslant 1) \\ +\infty, (\exists\ i,\ y_i(\overrightarrow{w} \cdot \overrightarrow{x_i} + b) < 1) \end{cases}$

target: $min_{\overrightarrow{w},b}\ max_{\alpha_i \geqslant 0}\ L(\overrightarrow{w}, b, \overrightarrow{\alpha}) = p^*$
利用拉格朗日函数对偶性，$max_{\alpha_i \geqslant 0}\ min_{\overrightarrow{w},b}\ L(\overrightarrow{w}, b, \overrightarrow{\alpha}) = d^*$
若要满足$p^* = d^*$，则需要满足**凸优化**和**KKT**条件。

solve

KKT:
$$\begin{cases} \alpha_i \geqslant 0 \\ y_i(\overrightarrow{w} \cdot \overrightarrow{x_i} + b) - 1 \geqslant 0 \\ \alpha_i(y_i(\overrightarrow{w} \cdot \overrightarrow{x_i} + b) - 1) = 0 \end{cases}$$

在满足**KKT**的情况下，易证**凸优化**成立。

求极值，需要满足
$$\begin{cases} \frac{\partial L}{\partial \overrightarrow{w}} = 0 = \overrightarrow{w} - \sum \alpha_i y_i \overrightarrow{x_i} \\ \frac{\partial L}{\partial b} = 0 = -\sum \alpha_i y_i \end{cases}$$
代入目标函数，有

$$
\begin{aligned}
L(\overrightarrow{w}, b, \overrightarrow{\alpha}) =& \frac{1}{2}(\sum \alpha_i y_i \overrightarrow{x_i})^2 - \sum \alpha_i \{y_i[(\sum \alpha_j y_j \overrightarrow{x_j}) \cdot \overrightarrow{x_i} + b] - 1\} \\
=& -\frac{1}{2}(\sum \alpha_i y_i \overrightarrow{x_i})^2 - b \sum \alpha_i y_i + \sum \alpha_i \\
=& -\frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j (\overrightarrow{x_i} \cdot \overrightarrow{x_j}) + \sum \alpha_i
\end{aligned}
$$

$max\ L = min\ -L$，用SMO求得$\overrightarrow{\alpha}$，则$\overrightarrow{w} = \sum \alpha_i y_i \overrightarrow{x_i}$，下面用反证法求$b$。
由**KKT**条件，有$\alpha_i(y_i(\overrightarrow{w} \cdot \overrightarrow{x_i} + b) - 1) = 0$。若$\forall \alpha_i = 0$，则$\overrightarrow{w} = 0$，矛盾，故$\exists \alpha_j \neq 0$，则解$y_i(\overrightarrow{w} \cdot \overrightarrow{x_i} + b) - 1 = 0$，可得$b = \frac{1}{y_i} - \overrightarrow{w} \cdot \overrightarrow{x_i}$。

soft margin

soft margin允许某些点不满足约束$y_i(\overrightarrow{w} \cdot \overrightarrow{x_i} + b) \geqslant 1$。
采用hinge损失，将原问题转化为$min_{\overrightarrow{w}, b, \overrightarrow{\xi}} \frac{1}{2}\overrightarrow{w}^2 + C \sum \xi_i$，满足$\begin{cases} y_i(\overrightarrow{w} \cdot \overrightarrow{x_i} + b) \geqslant 1 - \xi_i \\ \xi_i \geqslant 0 \end{cases}$
$\overrightarrow{\xi}$为松弛变量，$\xi_i = max\ (0,\ 1 - y_i(\overrightarrow{w} \cdot \overrightarrow{x_i} + b))$；$C > 0$称为惩罚函数。

$L(\overrightarrow{w}, b, \overrightarrow{\xi}, \overrightarrow{\alpha}, \overrightarrow{\mu}) = \frac{1}{2}\overrightarrow{w}^2 + C \sum \xi_i - \sum \alpha_i [y_i(\overrightarrow{w} \cdot \overrightarrow{x_i} + b) - (1 - \xi_i)] - \sum \mu_i \xi_i$
$\theta(\overrightarrow{w}, b, \overrightarrow{\xi}) = max_{\alpha_i \geqslant 0, \mu_i \geqslant 0}\ L(\overrightarrow{w}, b, \overrightarrow{\xi}, \overrightarrow{\alpha}, \overrightarrow{\mu}) =$
$\begin{cases} \frac{1}{2}\overrightarrow{w}^2 + C \sum \xi_i, (\forall i, \begin{cases} y_i(\overrightarrow{w} \cdot \overrightarrow{x_i} + b) - (1 - \xi_i) \geqslant 0 \\ \xi_i \geqslant 0 \end{cases} \\ +\infty, otherwise \end{cases}$

target: $min_{\overrightarrow{w}, b, \overrightarrow{\xi}}\ max_{\overrightarrow{\alpha} \geqslant 0, \overrightarrow{\mu} \geqslant 0}\ L(\overrightarrow{w}, b, \overrightarrow{\xi}, \overrightarrow{\alpha}, \overrightarrow{\mu}) = p^*$
$max_{\overrightarrow{\alpha} \geqslant 0, \overrightarrow{\mu} \geqslant 0}\ min_{\overrightarrow{w}, b, \overrightarrow{\xi}}\ L(\overrightarrow{w}, b, \overrightarrow{\xi}, \overrightarrow{\alpha}, \overrightarrow{\mu}) = d^*$

KKT:

$$
\begin{cases}
\alpha_i \geqslant 0 \\
y_i(\overrightarrow{w} \cdot \overrightarrow{x_i} + b) - (1 - \xi_i) \geqslant 0 \\
\alpha_i(y_i(\overrightarrow{w} \cdot \overrightarrow{x_i} + b) - (1 - \xi_i)) = 0 \\
\mu_i \geqslant 0 \\
\xi_i \geqslant 0 \\
\mu_i \xi_i = 0
\end{cases}
$$

$$
\begin{cases}
\frac{\partial L}{\partial \overrightarrow{w}} = 0 = \overrightarrow{w} - \sum \alpha_i y_i \overrightarrow{x_i} \\
\frac{\partial L}{\partial b} = 0 = -\sum \alpha_i y_i \\
\frac{\partial L}{\partial \overrightarrow{\xi}} = 0 = C - \overrightarrow{\alpha} - \overrightarrow{\mu}
\end{cases}
$$

代入得$L(\overrightarrow{w},b,\overrightarrow{\xi},\overrightarrow{\alpha},\overrightarrow{\mu})=-\frac{1}{2}\sum_{i,j}\alpha_i\alpha_jy_iy_j\overrightarrow{x_i}\cdot\overrightarrow{x_j}+\sum\alpha_i$

SMO求得$\overrightarrow{\alpha}$，则$\overrightarrow{w}=\sum\alpha_iy_i\overrightarrow{x_i}$。
对于soft margin而言，b是多解的。