# 支持向量机

支持向量机（Support Vector Machines，SVM）是一种监督学习算法，其主要用于二分类问题，同时也可以用于回归问题。本文使用支持向量机做为浅层学习器，其最终的决策函数只是由少量的支持向量确定的，因此计算的复杂度取决于支持向量的树木，而不是样本空间的位数，在某种意义上避免了由 One-Hot编码带来的“维数灾难”，因此适用于语义判断等高维场合，且正式由于其结果由少数支持向量决定，因此该模型具有较好的鲁棒性。但其分离超平面是通过二次规划求解，其中涉及矩阵运算，因此在大规模的训练样本上所需消耗的内存和运算时间都较长。

支持向量机的基本模型是定义在特征空间上的间隔最大化的线性分类器，间隔最大化学习策略本质上是一个求解凸二次规划（convex quadratic programming）的问题。支持向量机根据数据特征是否可分，可以分为线性可分支持向量机（硬间隔最大化）、线性支持向量机（软间隔最大化）以及非线性支持向量机（通过核函数将非线性转换为线性）。[1]

下面分别介绍线性可分支持向量机和线性支持向量机。

# 1. 线性可分支持向量机

对于线性可分的数据，支持向量机的基本思想是找出能够正确划分数据的间隔（margin）最大化的超平面，这里的间隔最大化指的是硬间隔最大化（hard margin maximization），即分离超平面中不允许有样本点出现。间隔最大化问题可以转换为其倒数的最小化问题（原始问题），并通过拉格朗日对偶性求解最小化规划问题的解（对偶问题），最后得到分离超平面以及分离决策函数来对新样本进行分类预测。

最大化间隔 相当于最小化 ：

为了求解最小化问题，对不等式约束条件

引入拉格朗日乘子（Largrange multiplier）:

定义拉格朗日函数：

其中， 为拉格朗日乘子向量。

根据拉格朗日对偶性，原始问题的对偶问题是极大极小问题：

所以为了得到对偶问题的解，需要先对 求 的极小，再对 求极大。

(1) 求

将拉格朗日函数 分别对 求偏导数，并令其等于0:

得：

将 (7) 代入拉格朗日公式 (4)，得：

即

（2）求 对 的极大，即使对偶问题：

将公式 (10) 的最大化转换为最小化，就得到了下面与之等价的对偶最优化问题：

设 是对偶最优化问题 (12-13) 的解，则存在下标 ，并可按下式求得原始最优化问题 (1) 的解 ：

由此得到分离超平面：

分离决策函数：

但时，对应的负向感情数据，当时，对应的正向感情数据，当时，可任意判断其情感。

# 2. 线性支持向量机

对线性不可分训练数据，上述方法中的不等式约束并不能满足函数间隔大于 1 的约束条件 (1)。为了解决这个问题，可以对每个样本点 引入一个松弛变量 ，使函数间隔加上松弛变量大于等于 1 。这样，约束条件变为：

我们称为软间隔最大化（soft margin maximization）。此时目标函数由原来的 变成：

其中 称为惩罚系数，当 值越大时对误分类的惩罚增大， 值小时对误分类的惩罚减小，近似线性可分的线性支持向量机的学习问题变成如下凸二次规划问题（原始问题）：

其拉格朗日函数为：

其中，。对偶问题是拉格朗日函数的极大极小问题，首先求 对 的极小并代入 (22) 得：

再对 求 的极大，即得对偶问题：

将对偶最优化问题 (24-25) 进行变换：利用 (25) 第二条公式消去 ，从而只留下变量 ，同时将目标函数求极大转换为求极小，可得原始问题 (39-40)的对偶问题 (26-27)：

设 是对偶问题 (26-27) 的一个解，若存在 的一个分量 ，，则原始问题 (20-21) 的解 可按下式求得：

由此得到分离超平面：

分离决策函数：

但时，对应的负向感情数据，当时，对应的正向感情数据，当时，可任意判断其情感。

# 参考资料

[1] 李航. 统计学习方法[M]. 北京: 清华大学出版社, 2012: 95-115.