

1.2.4. Complexity

Shinya Kunii

SVMはパワフルだが...

- 計算量とメモリ使用量

- Training dataの素性の数が増えることに増大

- SVMは、二次計画問題を解く

- 計算量は、

- $O(n_{features} * n_{samples}^2)$ と $O(n_{features} * n_{samples}^3)$

- の間

- もし、データがスパースなら、 $n_{features}$ をnon-vectorの素性の数の平均に置き換えができる。

- e.g. $n_{features} = 100000,$

- 一つの用例には、 $n_{features}$ の99.9%は0になっている

- $n_{features} \leftarrow 100000 * (1 - 0.999) = 100$

SVC と LinearSVC

- SVMでは、
 - カーネルに線形カーネルを利用するもの
 - 線形カーネル専用のSVMが存在する。
- LinearSVCの方が
 - かなり効果的に学習（並列計算機能）
 - 数百万の素性や用例を利用して、パラメータを直線的に調整できる利点を持っている。