Soft-Margin Support Vector Machine

Chih-Hao Chang

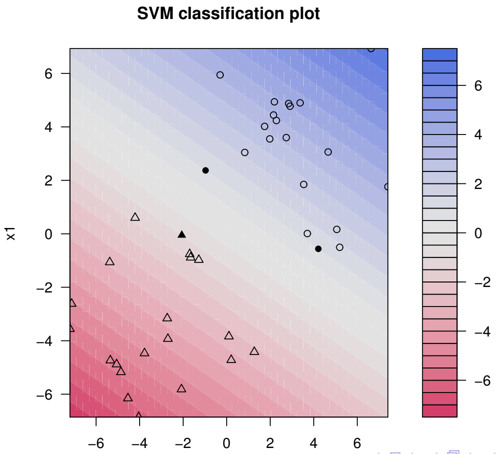
Institute of Statistics, National University of Kaohsiung
@ISS Academia Sinica

Aug. 02, 2018

1. Review for Hard-Margin SVM
2. Hard-Margin Versus Soft-Margin
3. Soft-Margin SVM
4. Discussion

Review on linear and kernel SVM

1. Data set: $\{y_n, \boldsymbol{x}_n\}_{n=1}^N$, $y_n \in \{+1, -1\}$ and $\boldsymbol{x}_n \in \mathbb{R}^d$.
2. Example for $d = 2$ and $N = 40$:



Review on linear and kernel SVM

1. Original problem of a linear classifier, $h(\boldsymbol{x}) = \boldsymbol{w}'\boldsymbol{x} + b = 0$:

$$\begin{aligned} \max \quad & \text{margin}(h), \\ \text{subject to} \quad & h \text{ classifies every } (\boldsymbol{x}_n, y_n) \text{ correctly}, \\ & \text{margin}(h) = \min_{n=1,\ldots,N} \text{distance}(\boldsymbol{x}_n, h). \end{aligned}$$

2. Parametrization of the problem:

$$\begin{aligned} \max \quad & \text{margin}(h), \\ \text{subject to} \quad & y_n(\boldsymbol{w}'\boldsymbol{x}_n + b) > 0; \quad n = 1, \ldots, N, \\ & \text{margin}(h) = \min_{n=1,\ldots,N} \frac{1}{\|\boldsymbol{w}\|} y_n(\boldsymbol{w}'\boldsymbol{x}_n + b). \end{aligned}$$

Review on linear and kernel SVM

1. The modified problem of the classifier, $h(\boldsymbol{x}) = \boldsymbol{w}'\boldsymbol{x} + b = 0$:

$$\begin{aligned} \max \quad & \text{margin}(h), \\ \text{subject to} \quad & y_n(\boldsymbol{w}'\boldsymbol{x}_n + b) > 0; \quad n = 1, \ldots, N, \\ & \text{margin}(h) = \min_{n=1,\ldots,N} \frac{1}{\|\boldsymbol{w}\|} y_n(\boldsymbol{w}'\boldsymbol{x}_n + b). \end{aligned}$$

2. Special scaling the minimal distance at 1:

$$\begin{aligned} \max_{b,\boldsymbol{w}} \quad & \frac{1}{\|\boldsymbol{w}\|}, \\ \text{subject to} \quad & \min_{n=1,\ldots,N} y_n(\boldsymbol{w}'\boldsymbol{x}_n + b) = 1. \end{aligned}$$

Review on linear and kernel SVM

1. Why can we scale the minimal distance at an arbitrary value?
   1. Suppose $\min_{n=1,\ldots,N} y_n(\boldsymbol{w}'\boldsymbol{x}_n + b) = 2$.
   2. The margin is then $2/\|\boldsymbol{w}\|$.
   3. To maximize $2/\|\boldsymbol{w}\|$ is the same to maximize $1/\|\boldsymbol{w}\|$.
   4. The optimization problem is

$$\max_{b,\boldsymbol{w}} \quad \frac{1}{\|\boldsymbol{w}\|},$$
$$\text{subject to} \quad \min_{n=1,\ldots,N} y_n\left(\frac{1}{2}\boldsymbol{w}'\boldsymbol{x}_n + \frac{b}{2}\right) = 1.$$

   and compare the the previous one:

$$\max_{b,\boldsymbol{w}} \quad \frac{1}{\|\boldsymbol{w}\|},$$
$$\text{subject to} \quad \min_{n=1,\ldots,N} y_n(\boldsymbol{w}'\boldsymbol{x}_n + b) = 1.$$

2. Obviously, $\boldsymbol{w} = \frac{1}{2}\boldsymbol{w} \Rightarrow 1/\|\boldsymbol{w}\| = 2/\|\boldsymbol{w}\|$.

Review on linear and kernel SVM

1. The scaled problem of the classifier, $h(\boldsymbol{x}) = \boldsymbol{w}'\boldsymbol{x} + b = 0$:

$$\max_{b,\boldsymbol{w}} \quad \frac{1}{\|\boldsymbol{w}\|},$$
$$\text{subject to} \quad \min_{n=1,\ldots,N} y_n(\boldsymbol{w}'\boldsymbol{x}_n + b) = 1.$$

2. Release of constraints:

$$\min_{b,\boldsymbol{w}} \quad \frac{1}{2}\boldsymbol{w}'\boldsymbol{w},$$
$$\text{subject to} \quad y_n(\boldsymbol{w}'\boldsymbol{x}_n + b) \geq 1; \quad n = 1, \ldots, N.$$

Review on linear and kernel SVM

1. Why can we release the constraints?
   1. If the optimal $(b, \boldsymbol{w}) \notin \{(b, \boldsymbol{w}) : \min_{n=1,\ldots,N} y_n(\boldsymbol{w}'\boldsymbol{x}_n + b) = 1\}$, then say $\min_{n=1,\ldots,N} y_n(\boldsymbol{w}'\boldsymbol{x}_n + b) = 1.126$.
   2. As we previously mentioned, there is another equivalent optimal pair: $(b, \boldsymbol{w}) = (b/1.126, \boldsymbol{w}/1.126)$.
   3. The resulting margin is

$$\frac{1}{\|\boldsymbol{w}\|} = \frac{1.126}{\|\boldsymbol{w}\|} > \frac{1}{\|\boldsymbol{w}\|},$$

   which fails the optimality of $(b, \boldsymbol{w})$, a contradiction.
2. Hence the optimal $(b, \boldsymbol{w}) \in \{(b, \boldsymbol{w}) : \min_{n=1,\ldots,N} y_n(\boldsymbol{w}'\boldsymbol{x}_n + b) = 1\}$

Review on linear and kernel SVM

1. The released problem:

$$\begin{aligned} \min_{b,\boldsymbol{w}} \quad & \frac{1}{2}\boldsymbol{w}'\boldsymbol{w}, \\ \text{subject to} \quad & y_n(\boldsymbol{w}'\boldsymbol{x}_n + b) \geq 1; \quad n = 1, \ldots, N. \end{aligned}$$

2. Solution from the process of quadratic programming (QP):

$$\begin{aligned} \min_{\boldsymbol{u}} \quad & \frac{1}{2}\boldsymbol{u}'\boldsymbol{Q}\boldsymbol{u} + \boldsymbol{p}'\boldsymbol{u}, \\ \text{subject to} \quad & \boldsymbol{a}_m\boldsymbol{u} \geq c_m; \quad m = 1, \ldots, M. \end{aligned}$$

Review on linear and kernel SVM

1. The linear SVM:

$$\min_{b,\boldsymbol{w}} \quad \frac{1}{2}\boldsymbol{w}'\boldsymbol{w}$$

$$\text{subject to} \quad y_n(\boldsymbol{w}'\boldsymbol{x}_n + b) \geq 1; \quad n = 1, \ldots, N,$$

2. Solution from the method of Lagrange multipliers:

$$\mathcal{L}(b, \boldsymbol{w}, \boldsymbol{\alpha}) = \frac{1}{2}\boldsymbol{w}'\boldsymbol{w} + \sum_{n=1}^{N} \alpha_n(1 - y_n(\boldsymbol{w}'\boldsymbol{x}_n + b)),$$

where

$$\text{SVM} = \min_{b,\boldsymbol{w}} \left(\max_{\alpha_n \geq 0} \mathcal{L}(b, \boldsymbol{w}, \boldsymbol{\alpha}) \right)$$

Review on linear and kernel SVM

1. Why

$$\text{SVM} = \min_{b,\boldsymbol{w}} \left(\max_{\alpha_n \geq 0} \left(\frac{1}{2}\boldsymbol{w}'\boldsymbol{w} + \sum_{n=1}^{N} \alpha_n (1 - y_n(\boldsymbol{w}'\boldsymbol{x}_n + b)) \right) \right)?$$

   1. Can $y_n(\boldsymbol{w}'\boldsymbol{x}_n + b) < 1$ be happened?
   2. If yes, $1 - y_n(\boldsymbol{w}'\boldsymbol{x}_n + b) > 0$, then $\alpha_n = \infty$.
   3. This cannot be a solution of SVM.
2. If $1 - y_n(\boldsymbol{w}'\boldsymbol{x}_n + b) < 0$, $\alpha_n = 0$.
3. Hence $\alpha_n > 0$ can only be happened on $1 - y_n(\boldsymbol{w}'\boldsymbol{x}_n + b) = 0$ (complementary slackness).

Review on linear and kernel SVM

1. For

$$\mathcal{L}(b, \boldsymbol{w}, \boldsymbol{\alpha}) = \frac{1}{2}\boldsymbol{w}'\boldsymbol{w} + \sum_{n=1}^{N} \alpha_n (1 - y_n(\boldsymbol{w}'\boldsymbol{x}_n + b)),$$

   the Lagrange dual problem (weak duality) is:

$$\text{SVM} = \min_{b,\boldsymbol{w}} \left(\max_{\alpha_n \geq 0} \mathcal{L}(b, \boldsymbol{w}, \boldsymbol{\alpha}) \right) \geq \max_{\alpha_n \geq 0} \left(\min_{b,\boldsymbol{w}} \mathcal{L}(b, \boldsymbol{w}, \boldsymbol{\alpha}) \right).$$

2. The equality holds for strong duality:
   1. Convex;
   2. Existence of he solution;
   3. Linear constraints.

Review on linear and kernel SVM

1. Simplify the dual problem: under the strong duality,

$$\text{SVM} = \max_{\alpha_n \geq 0} \left(\min_{b,\boldsymbol{w}} \frac{1}{2}\boldsymbol{w}'\boldsymbol{w} + \sum_{n=1}^{N} \alpha_n (1 - y_n(\boldsymbol{w}'\boldsymbol{x}_n + b)) \right)$$

$$= \max_{\alpha_n \geq 0, \sum y_n \alpha_n = 0} \left(\min_{\boldsymbol{w}} \frac{1}{2}\boldsymbol{w}'\boldsymbol{w} + \sum_{n=1}^{N} \alpha_n (1 - y_n(\boldsymbol{w}'\boldsymbol{x}_n)) \right),$$

since

$$\frac{\partial \mathcal{L}(b, \boldsymbol{w}, \boldsymbol{\alpha})}{\partial b} = -\sum_{n=1}^{N} \alpha_n y_n = 0.$$

Review on linear and kernel SVM

1. Further simplify the dual problem: under the strong duality and by

$$\frac{\partial \mathcal{L}(b, \boldsymbol{w}, \boldsymbol{\alpha})}{\partial w_i} = w_i - \sum_{n=1}^{N} \alpha_n y_n x_{n,i} = 0 \Rightarrow \quad \boldsymbol{w} = \sum_{n=1}^{N} \alpha_n y_n \boldsymbol{x}_n,$$

and hence

$$\begin{aligned} \text{SVM} &= \max_{\alpha_n \geq 0, \sum y_n \alpha_n = 0} \left(\min_{\boldsymbol{w}} \frac{1}{2} \boldsymbol{w}' \boldsymbol{w} + \sum_{n=1}^{N} \alpha_n (1 - y_n (\boldsymbol{x_n}' \boldsymbol{w})) \right) \\ &= \max_{\alpha_n \geq 0, \sum y_n \alpha_n = 0, \boldsymbol{w} = \sum \alpha_n y_n \boldsymbol{x}_n} \left(\frac{1}{2} \boldsymbol{w}' \boldsymbol{w} + \sum_{n=1}^{N} \alpha_n - \boldsymbol{w}' \boldsymbol{w} \right) \\ &= \max_{\alpha_n \geq 0, \sum y_n \alpha_n = 0, \boldsymbol{w} = \sum \alpha_n y_n \boldsymbol{x}_n} \left(-\frac{1}{2} \left\| \sum_{n=1}^{N} \alpha_n y_n \boldsymbol{x}_n \right\|^2 + \sum_{n=1}^{N} \alpha_n \right). \end{aligned}$$

Review on linear and kernel SVM

1. Karush-Kuhn-Tucker (KKT) conditions to solve b and $\boldsymbol{w}$ from optimal $\boldsymbol{\alpha}$:
    1. Primal feasible:
    $$y_n(\boldsymbol{w}'\boldsymbol{x}_n + b) \geq 1;$$
    2. Dual feasible:
    $$\alpha_n \geq 0;$$
    3. Dual-inner optimal:
    $$\sum y_n \alpha_n = 0, \text{ and } \boldsymbol{w} = \sum \alpha_n y_n \boldsymbol{x}_n;$$
    4. Primal-inner optimal (complementary slackness):
    $$\alpha_n(1 - y_n(\boldsymbol{w}'\boldsymbol{x}_n + b)) = 0.$$

Review on linear and kernel SVM

1. The dual problem:

$$\max_{\alpha_n \geq 0, \sum y_n \alpha_n = 0, \boldsymbol{w} = \sum \alpha_n y_n \boldsymbol{x}_n} \left(-\frac{1}{2} \left\| \sum_{n=1}^{N} \alpha_n y_n \boldsymbol{x}_n \right\|^2 + \sum_{n=1}^{N} \alpha_n \right)$$

2. The equivalent standard dual SVM problem:

$$\begin{aligned} \min_{\boldsymbol{\alpha}} \quad & \frac{1}{2} \sum_{n=1}^{N} \sum_{m=1}^{N} \alpha_n \alpha_m y_n y_m \boldsymbol{x}_n' \boldsymbol{x}_m - \sum_{n=1}^{N} \alpha_n, \\ \text{subject to} \quad & \sum_{n=1}^{N} y_n \alpha_n = 0; \\ & \alpha_n \geq 0; \quad \text{for } n = 1, \ldots, N, \end{aligned}$$

which has N variables and $N+1$ constraints, and is a convex quadratic programming problem.

Review on linear and kernel SVM

1. By KKT conditions, we have $\boldsymbol{w} = \sum \alpha_n y_n \boldsymbol{x}_n$ and for complementary slackness condition: $\alpha_n(1 - y_n(\boldsymbol{w}'\boldsymbol{x}_n + b)) = 0$; $n = 1, \ldots, N$,

$$b = y_s - \boldsymbol{w}'\boldsymbol{x}_s, \quad \text{with } \alpha_s > 0,$$

2. Or let $\mathcal{I} = \{s = 1, \ldots, N : \alpha_s > 0\}$,

$$b = \frac{1}{|\mathcal{I}|} \sum_{s \in \mathcal{I}} y_s - \boldsymbol{w}'\boldsymbol{x}_s,$$

where $\mathcal{I}$ is the set of support vectors: $y_s(\boldsymbol{w}'\boldsymbol{x}_n + b) = 1$; $s \in \mathcal{I}$.

Review on linear and kernel SVM

1. Primal linear SVM (suitable for small d):

$$\min_{b,\boldsymbol{w}} \quad \frac{1}{2}\boldsymbol{w}'\boldsymbol{w}$$
$$\text{subject to} \quad y_n(\boldsymbol{w}'\boldsymbol{x}_n + b) \geq 1; \quad n = 1, \ldots, N.$$

   1. $d + 1$ variables;
   2. N constraints;

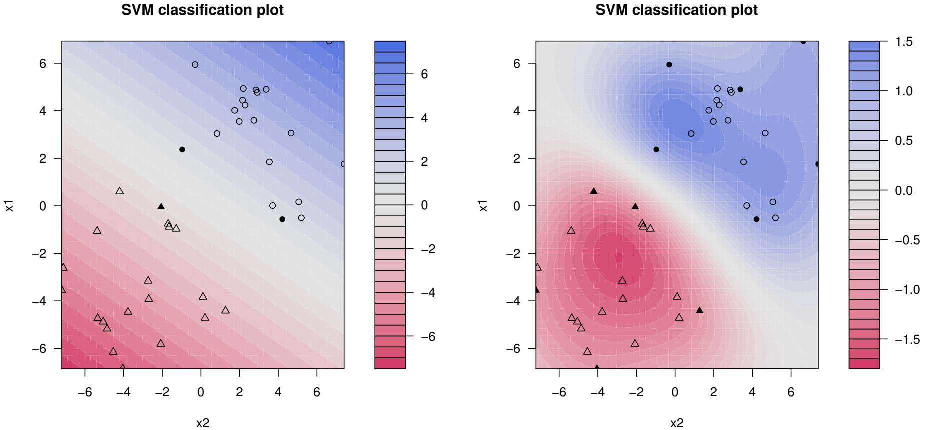
2. Dual SVM (suitable for small N):

$$\min_{\boldsymbol{\alpha}} \quad \frac{1}{2}\boldsymbol{\alpha}'\boldsymbol{Q}\boldsymbol{\alpha} - \mathbf{1}'\boldsymbol{\alpha}; \quad \boldsymbol{Q} = \{q_{n,m}\};\ q_{n,m} = y_n y_m \boldsymbol{x}_n' \boldsymbol{x}_m,$$
$$\text{subject to} \quad \boldsymbol{y}'\boldsymbol{\alpha} = 0;$$
$$\alpha_n \geq 0; \quad \text{for } n = 1, \ldots, N,$$

   1. N variables;
   2. $N + 1$ constraints.

Review on linear and Kernel SVM

1. How about non-linear boundary?



Review on linear and kernel SVM

1. Let $\Phi : \mathbb{R}^d \to \mathbb{R}^{\tilde{d}}$ and $\boldsymbol{z}_n = \Phi(\boldsymbol{x}_n)$; $n = 1, \ldots, N$.
2. Kernel function:

$$K_\Phi(\boldsymbol{x}_n, \boldsymbol{x}_m) = \Phi(\boldsymbol{x}_n)'\Phi(\boldsymbol{x}_m).$$

3. The dual problem on the transform function Φ:

$$\begin{aligned} \min_{\boldsymbol{\alpha}} \quad & \frac{1}{2}\boldsymbol{\alpha}'\boldsymbol{Q}_\Phi\boldsymbol{\alpha} - \mathbf{1}'\boldsymbol{\alpha};, \\ \text{subject to} \quad & \boldsymbol{y}'\boldsymbol{\alpha} = 0; \\ & \alpha_n \geq 0; \quad \text{for } n = 1, \ldots, N, \end{aligned}$$

where $\boldsymbol{Q}_\Phi = \{q_{n,m}\}$; $q_{n,m} = y_n y_m K_\Phi(\boldsymbol{x}_n, \boldsymbol{x}_m)$.

Review on linear and kernel SVM

1. Polynomial kernel function:

$$K_q(\boldsymbol{x}, \boldsymbol{x}^*) = (\zeta + \gamma \boldsymbol{x}' \boldsymbol{x}^*)^q; \quad \gamma > 0,\ \zeta \geq 0,$$

   which is commonly applied for polynomial SVM.

2. Gaussian kernel function:

$$K(\boldsymbol{x}, \boldsymbol{x}^*) = \exp(-\gamma \|\boldsymbol{x} - \boldsymbol{x}^*\|^2), \quad \gamma > 0,$$

   which is a kernel of infinite dimensional transform and for $d = 1$ (i.e. $\boldsymbol{x} = x$),

$$\Phi(x) = \exp(-x^2) \cdot \left(1, \sqrt{\frac{2}{1!}}x, \sqrt{\frac{2^2}{2!}}x^2, \dots\right),$$

Review on linear and kernel SVM

1. The classifier of kernel svm, $g_{\text{svm}}(\boldsymbol{x}) = \text{sign}(\boldsymbol{w}'\Phi(\boldsymbol{x}) + b)$:

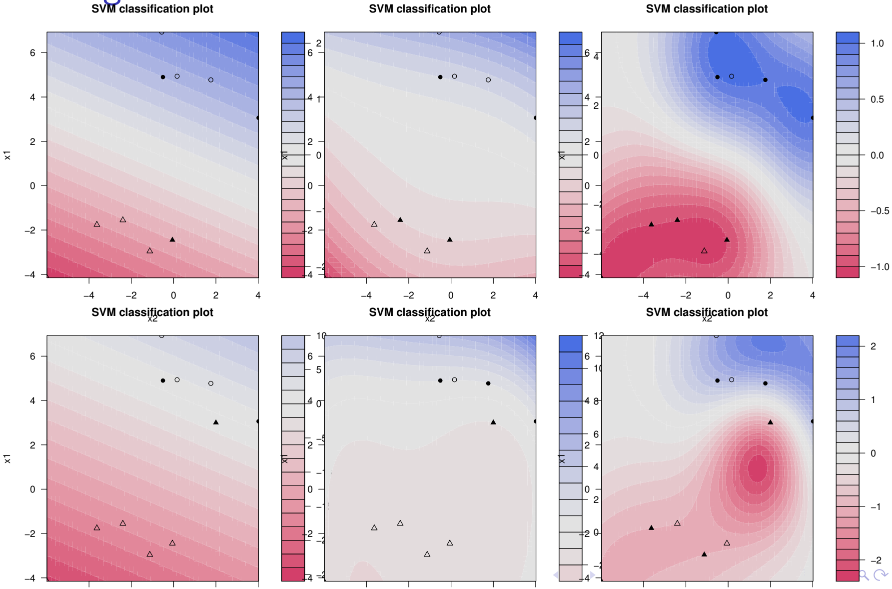
$$b = y_s - \boldsymbol{w}'\boldsymbol{z}_s = y_s - \Big(\sum_{n=1}^{N} \alpha_n y_n \boldsymbol{z}_n\Big)' \boldsymbol{z}_s$$

$$= y_s - \sum_{n=1}^{N} \alpha_n y_n K(\boldsymbol{x}_n, \boldsymbol{x}_s),$$

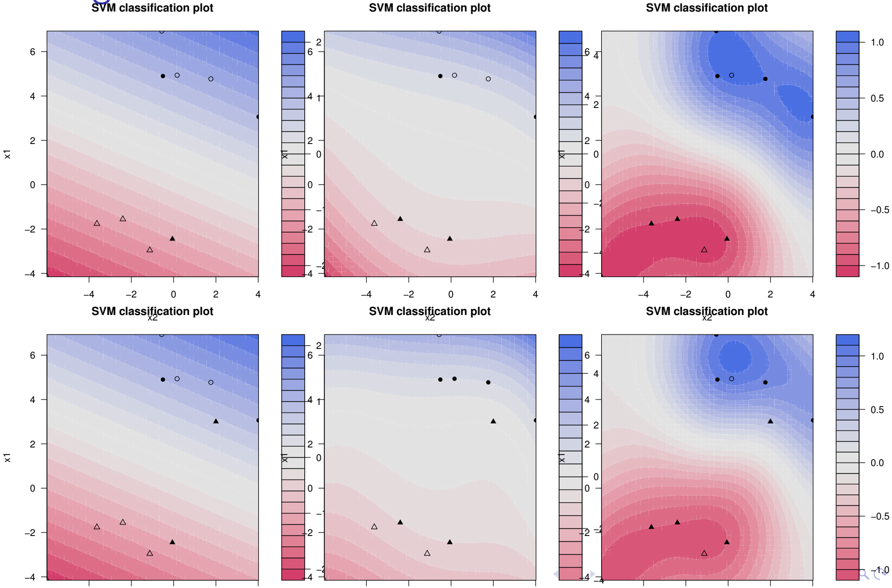
$$\boldsymbol{w}'\Phi(\boldsymbol{x}) = \sum_{n=1}^{N} \alpha_n y_n \boldsymbol{z}_n' \boldsymbol{z} = \sum_{n=1}^{N} \alpha_n y_n K(\boldsymbol{x}_n, \boldsymbol{x}).$$

2. We don't need to know $\boldsymbol{w}$ and the hyperplane classifier can be a mystery.

Hard-Margin SVM



Soft-Margin SVM



Soft-Margin SVM

1. If always insisting on separable, SVM tends overfitting to noise.
2. Hard-margin SVM:

$$\min_{b,\boldsymbol{w},\boldsymbol{\xi}} \quad \frac{1}{2}\boldsymbol{w}'\boldsymbol{w},$$
$$\text{subject to} \quad y_n(\boldsymbol{w}'\boldsymbol{x}_n + b) \geq 1.$$

3. Tolerance on the noise by allowing the data points being incorrectly classified.
4. Record the margin violation by ξ_n; $n = 1, \ldots, N$.
5. Soft-margin SVM:

$$\min_{b,\boldsymbol{w},\boldsymbol{\xi}} \quad \frac{1}{2}\boldsymbol{w}'\boldsymbol{w} + C\sum_{n=1}^{N}\xi_n$$
$$\text{subject to} \quad y_n(\boldsymbol{w}'\boldsymbol{x}_n + b) \geq 1 - \xi_n,$$
$$\xi_n \geq 0; \quad n = 1, \ldots, N,$$

Soft-Margin Dual SVM

1. Quadratic programming with $d + 1 + N$ variables and $2N$ constraints:

$$\min_{b,\boldsymbol{w},\boldsymbol{\xi}} \quad \frac{1}{2}\boldsymbol{w}'\boldsymbol{w} + C\sum_{n=1}^{N}\xi_n$$

$$\text{subject to} \quad y_n(\boldsymbol{w}'\boldsymbol{x}_n + b) \geq 1 - \xi_n,$$

$$\xi_n \geq 0; \quad n = 1, \ldots, N,$$

2. The dual problem: $\text{SVM} = \min_{b,\boldsymbol{w},\boldsymbol{\xi}} \left(\max_{\alpha_n \geq 0, \beta_n \geq 0} \mathcal{L}(b, \boldsymbol{w}, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\beta}) \right)$,

$$\mathcal{L}(b, \boldsymbol{w}, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \frac{1}{2}\boldsymbol{w}'\boldsymbol{w} + C\sum_{n=1}^{N}\xi_n + \sum_{n=1}^{N}\beta_n(-\xi_n)$$

$$+ \sum_{n=1}^{N}\alpha_n(1 - \xi_n - y_n(\boldsymbol{w}'\boldsymbol{z}_n + b)).$$

Soft-Margin Dual SVM

1. Simplify the soft-margin dual problem: under the strong duality,

$$\text{SVM} = \max_{\alpha_n \geq 0, \beta_n \geq 0} \Bigg(\min_{b, \boldsymbol{w}, \boldsymbol{\xi}} \frac{1}{2}\boldsymbol{w}'\boldsymbol{w} + C\sum_{n=1}^{N} \xi_n + \sum_{n=1}^{N} \beta_n(-\xi_n) + \sum_{n=1}^{N} \alpha_n (1 - \xi_n - y_n(\boldsymbol{w}'\boldsymbol{z}_n + b)) \Bigg)$$

$$= \max_{0 \leq \alpha_n \leq C, \beta_n = C - \alpha_n} \Bigg(\min_{b, \boldsymbol{w}} \frac{1}{2}\boldsymbol{w}'\boldsymbol{w} + \sum_{n=1}^{N} \alpha_n (1 - y_n(\boldsymbol{w}'\boldsymbol{z}_n + b)) \Bigg),$$

since

$$\frac{\partial \mathcal{L}}{\partial \xi_n} = C - \alpha_n - \beta_n = 0 \Rightarrow \beta_n = C - \alpha_n \geq 0.$$

Soft-Margin Dual SVM

1. Further simplify the soft-margin dual problem: under the strong duality and by

$$\frac{\partial \mathcal{L}}{\partial b} = 0 \;\Rightarrow\; \sum_{n=1}^{N} \alpha_n y_n = 0,$$

$$\frac{\partial \mathcal{L}}{\partial w_i} = 0 \;\Rightarrow\; \boldsymbol{w} = \sum_{n=1}^{N} \alpha_n y_n \boldsymbol{z}_n,$$

we have

$$\text{SVM} = \max_{0\le\alpha_n\le C,\beta_n=C-\alpha_n,\sum y_n\alpha_n=0,\boldsymbol{w}=\sum\alpha_n y_n \boldsymbol{z}_n} \left\{ -\frac{1}{2}\|\sum_{n=1}^{N}\alpha_n y_n \boldsymbol{z}_n\|^2 + \sum_{n=1}^{N}\alpha_n \right\}.$$

Soft-Margin Dual SVM

1. The soft-margin dual SVM:

$$\text{SVM} = \max_{0\le\alpha_n\le C, \beta_n = C-\alpha_n, \sum y_n\alpha_n = 0, \boldsymbol{w}=\sum \alpha_n y_n \boldsymbol{z}_n} \Big\{ -\frac{1}{2}\|\sum_{n=1}^{N}\alpha_n y_n \boldsymbol{z}_n\|^2 + \sum_{n=1}^{N}\alpha_n \Big\}.$$

2. QP with N variables and $2N+1$ constraints:

$$\min_{\boldsymbol{\alpha}} \quad \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N}\alpha_n\alpha_m y_n y_m \boldsymbol{z}_n'\boldsymbol{z}_m - \sum_{n=1}^{N}\alpha_n,$$

$$\text{subject to} \quad \sum_{n=1}^{N} y_n\alpha_n = 0,\ 0 \le \alpha_n \le C;\ n = 1,\dots,N,$$

$$\text{implicity} \quad \boldsymbol{w} = \sum_{n=1}^{N}\alpha_n y_n \boldsymbol{z}_n,\ \beta_n = C - \alpha_n;\ n = 1,\dots,N,$$

Soft-Margin Kernel SVM

1. QP with N variables and $2N+1$ constraints:

$$\min_{\boldsymbol{\alpha}} \quad \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N}\alpha_n\alpha_m y_n y_m K_\Phi(\boldsymbol{x}_n, \boldsymbol{x}_m) - \sum_{n=1}^{N}\alpha_n,$$

$$\text{subject to} \quad \sum_{n=1}^{N} y_n\alpha_n = 0,\ 0 \le \alpha_n \le C;\ n = 1, \ldots, N,$$

$$\text{implicity} \quad \boldsymbol{w} = \sum_{n=1}^{N}\alpha_n y_n \boldsymbol{z}_n,\ \beta_n = C - \alpha_n;\ n = 1, \ldots, N,$$

2. The hypothesis hyperplane of soft-margin kernel svm:

$$g_{\text{svm}}(\boldsymbol{x}) = \text{sign}\Big(\sum_{n=1}^{N}\alpha_n y_n K_\Phi(\boldsymbol{x}_n, \boldsymbol{x}) + b\Big),$$

where b can be solved by the complementary slackness.

Soft-Margin Kernel SVM

1. The hypothesis hyperplane:

$$g_{\text{svm}}(\boldsymbol{x}) = \text{sign}\left(\sum_{n=1}^{N} \alpha_n y_n K_\Phi(\boldsymbol{x}_n, \boldsymbol{x}) + b\right),$$

2. The complementary slackness conditions:

$$\alpha_n(1 - \xi_n - y_n(\boldsymbol{w}'\boldsymbol{z}_n + b)) = 0,$$
$$(C - \alpha_n)\xi_n = 0.$$

   1. For support vectors (with $\alpha_s > 0$):

$$b = y_s - y_s\xi_s - \boldsymbol{w}'\boldsymbol{z}_s.$$

   2. For free support vectors (with $0 < \alpha_s < C$):

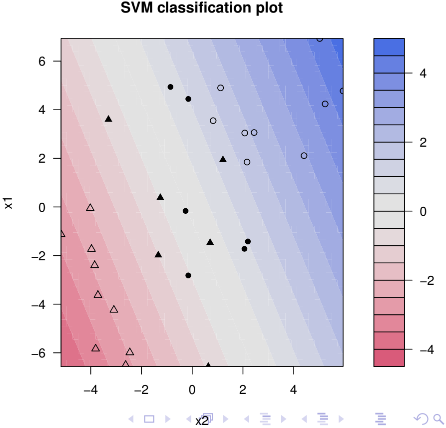
$$\xi_s = 0 \Rightarrow b = y_s - \boldsymbol{w}'\boldsymbol{z}_s.$$

Physical Meanings of α_n

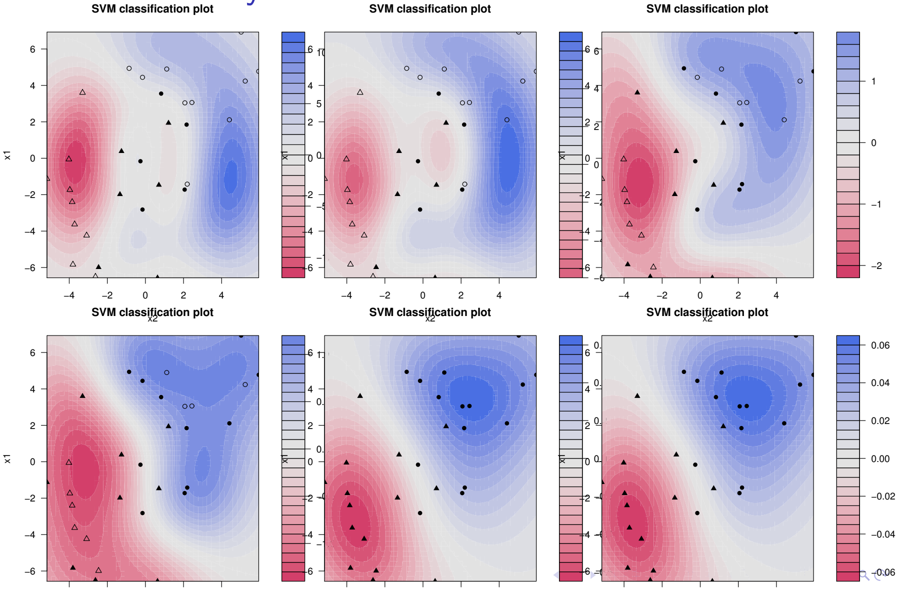
1. Complementary slackness:

$$\alpha_n(1 - \xi_n - y_n(\boldsymbol{w}'\boldsymbol{z}_n + b)) = 0,$$
$$(C - \alpha_n)\xi_n = 0.$$

1. Non-SV ($\alpha_n = 0$): $\xi_n = 0$;
2. SV ($0 < \alpha_n < C$): $\xi_n = 0$;
3. Bounded SV ($\alpha_n = C$): $\xi_n = 1 - y_n(\boldsymbol{w}'\boldsymbol{z}_n + b) \geq 0$.
4. α_n can be used for data analysis.



Selection of Penalty C



R-code of the toy example

```
library(kernlab)
set.seed(500)
n=10
# data generation
x1 = c(rnorm(n,3,2),rnorm(n,-3,2),rnorm(n,0,2))
x2 = c(rnorm(n,3,2),rnorm(n,-3,2),rnorm(n,0,2))
y = factor(c(rep(T,n),rep(F,n),rep(c(T,F),n/2)))
data = data.frame(y =y,x1=x1,x2=x2)
# fit the soft-margin Gaussian kernel SVM
model.ksvm = ksvm(y ~ x1 + x2, data = data, kernel="rbfdot",
    kpar=list(sigma=1),C=1)
plot(model.ksvm, data=data)
```

References

1. Special thanks to Prof. Hsuan-Tien Lin
   1. Handout slides and youtube vedios: `https://www.csie.ntu.edu.tw/~htlin/mooc/`