Support Vector machine
with Random variables

Professor Abdellatif El Afia

Plan

1. Robust Support Vector Classification
2. Robust Support Vector Regression
3. Probabilistic Constraints Support Vector Classification
4. Least Squares Probabilistic Support Vector Classification
5. Probabilistic Constraints Support Vector Regression

Robust Support vector machine

Uncertain Data

$$x_i = \overline{x_i} + \Delta x_i$$

Robust Support Vector Classification

1. Traditional Soft Margin SVC
2. Total Support Vector Classification (TSVC)
3. Robust SVC with Bounded Uncertainty
4. Robust SVC with Ellipsoidal Uncertainty
5. Robust SVC with Polyhedral Uncertainty
6. Uncertainty Support Vector Classification (USVC)

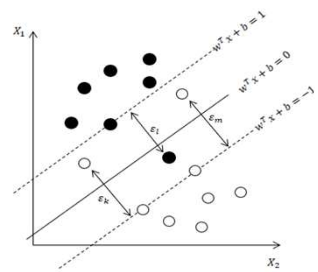
Traditional Soft margin SVC (C-SVC)

Suppose we have a two-class dataset of m data points $\{x_i, y_i\}_{\{i=1\}}^{m}$ with n $-$dimensional features $x_i \in \mathbb{R}^n$ and respective class labels $y_i \in \{+1, -1\}$. The soft margin SVC formulation is :

$$\min_{w,b,\xi_i} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m} \xi_i$$

$$s.t.\ \ y_i(w^T x_i + b) \geq 1 - \xi_i, \qquad i = 1, \dots, m,$$

$$\xi_i \geq 0, \qquad i = 1, \dots, m$$



Where w is the weight of the optimal classifying hyperplane $w^T x + b = 0$, b is the bias, C is a trade-off parameter between the maximum margin and the errors, and ξ_i is the slack variable.

Uncertain Data: Examples

The uncertain data can be used to represent :

- The range of possible arrival times for flights.
- The range of possible values for sensor measurements.
- The range of possible values for patient features such as blood pressure and heart rate for medical diagnosis.

Total Support Vector Classification (TSVC)

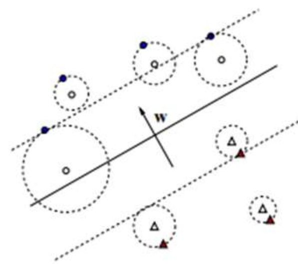
Assuming the data points are subject to an additive noise, $x_i = \overline{x_i} + \Delta x_i$ and the noise is bounded by$\|\Delta x_i\| \leq \delta_i$. Then the problem above is redefined as follows:

$$\min_{w,b,\xi_i} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m} \xi_i$$

$$s.t.\ \ y_i(w^T(\overline{x_i} + \Delta x_i) + b) \geq 1 - \xi_i, \qquad i = 1, \dots, m,$$

$$\xi_i \geq 0, \qquad i = 1, \dots, m,$$

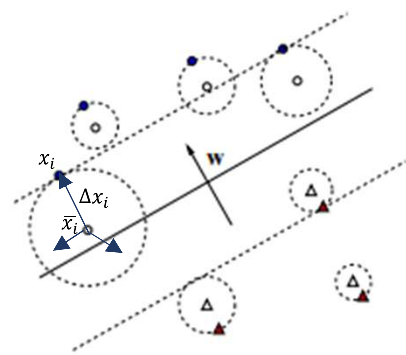
$$\|\Delta x_i\| \leq \delta_i, \qquad i = 1, \dots, m$$



The uncertain data x_i can be represented by the circle centered at $\overline{x_i}$ with radius equal to δ_i.

Total Support Vector Classification (TSVC)

The problem here is by assuming that $x_i = \overline{x_i} + \Delta x_i$, x_i could move toward any direction in the uncertainty set.



Robust SVC with Bounded Uncertainty

Therefore we consider w as a robust feasible solution **if and only if** for every i, $i = 1, \ldots, m$, we have:

$$\min_{\|\Delta x_i\| \leq \delta_i} y_i(w^T(\overline{x_i} + \Delta x_i) + b) \geq 1 - \xi_i$$

First, we need to solve the following problem (in which w is fixed):

$$\min_{\Delta x_i} y_i w^T \Delta x_i$$
$$s.t. \|\Delta x_i\| \leq \delta_i, \qquad i = 1, \ldots, m$$

Robust SVC with Bounded Uncertainty

According to Hölder's inequality, we have :

$$|y_i w^T \Delta x_i| \leq \|\Delta x_i\| \cdot \|w\| \leq \delta_i \|w\|$$

Thus, a lower bound of $\boldsymbol{y_i w^T} \Delta \boldsymbol{x_i}$ is $-\boldsymbol{\delta_i}\|\boldsymbol{w}\|$,
And by substituting into the original problem we get the following formulation:

$$\min_{w,b,\xi_i} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m} \xi_i$$

$$s.t. \;\; y_i(w^T \overline{x_i} + b) - \delta_i \|w\| \geq 1 - \xi_i, \qquad i = 1, \ldots, m,$$

$$\xi_i \geq 0, \qquad i = 1, \ldots, m$$

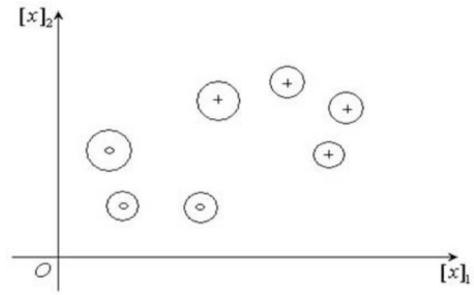
Robust SVC with Elipsoidal Uncertainty

Now the direction of how the data points are perturbed is not important because we will work with the boundaries of the Super-sphere set, where the input data is defined by: $x_i \in \{x \mid \| x - \overline{x_i} \| \leq \delta_i\}$

In the case of $\mathbb{R}^2$ the uncertainty set is represented by circles with center $\overline{x_i}$ and radius δ_i.

Let $\overline{x_i} = (1, 2)$, and $\delta_i = 2$. This super-sphere is a circle defined by:

$$([x]_1 - 1)^2 + ([x]_2 - 2)^2 \leq 2^2$$



Robust SVC with Elipsoidal Uncertainty

Theorem:

The triple $(w^*, b^*, \xi^*)$is a solution to the primal problem if and only if it is a solution to the dual form w.r.t. (w, b, ξ).

In order to convert the above problem into the dual form for a simpler solution, we first introduce a variable t with $\|w\| \leq t$, and write is as :

$$\min_{w,b,\xi_i} \frac{1}{2} t^2 + C \sum_{i=1}^{m} \xi_i$$

$$s.t.\ \ y_i(w^T \overline{x_i} + b) - \delta_i t \geq 1 - \xi_i, \qquad i = 1, \dots, m,$$

$$\xi_i \geq 0, \qquad i = 1, \dots, m,$$

$$\|w\| \leq t$$

Robust SVC with Elipsoidal Uncertainty

Furthermore, introduce two variables u and v with the constraints $u + v = 1$ and $\sqrt{t^2 + v^2} \leq u$. Therefore, we have $t^2 = u^2 - v^2 = (u - v)(u + v) = u - v$.
By replacing t^2 by $u - v$ and rewriting the constraints in the previous problem the following problem can be obtained:

$$\min_{w,b,\xi_i} \frac{1}{2}(u - v) + C\sum_{i=1}^{m} \xi_i$$

$$s.t.\ \ y_i(w^T\overline{x_i} + b) - \delta_i t \geq 1 - \xi_i, \qquad i = 1, \dots, m,$$

$$\xi_i \geq 0, \qquad i = 1, \dots, m,$$

$$u + v = 1,$$

$$\sqrt{t^2 + v^2} \leq u,$$

$$\|w\| \leq t$$

Robust SVC with Ellipsoidal Uncertainty

The Lagrange function for this problem is as follows:

$$L = \frac{1}{2}(u - v) + C\sum_{i=1}^{m} \xi_i - \sum_{i=1}^{m} \alpha_i (y_i (w^T \overline{x_i} + b) - \delta_i t - 1 + \xi_i) - \sum_{i=1}^{m} \eta_i \xi_i$$
$$- \beta(u + v - 1) - z_u u - z_v v - \gamma t - z_t t - z_w^T w,$$

Where $\alpha, \eta \in \mathbb{R}^m$, $\beta, z_u, z_v, \gamma, z_t \in \mathbb{R}$, $z_w \in \mathbb{R}^n$ are the multiplier vectors.

Robust SVC with Ellipsoidal Uncertainty

Thus, the dual problem can be obtained:

$$\max_{\alpha,\beta,\gamma,z_u,z_v} \beta + \sum_{i=1}^{m} \alpha_i ,$$

$$s.t. \quad \gamma \le \beta + \sum_{i=1}^{m} \delta_i \alpha_i - \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j \left(x_i{}^T x_j\right)},$$

$$\beta + z_u = \frac{1}{2}, \qquad \beta + z_v = \frac{1}{2},$$

$$\sum_{i=1}^{m} y_i \alpha_i = 0,$$

$$0 \le \alpha_i \le C, i = 1, \cdots, m,$$

$$\sqrt{\gamma^2 + z_v^2} \le z_u.$$

Robust SVC with Ellipsoidal Uncertainty

Theorem:

Suppose that $\left(\alpha^{*T}, \gamma^*\right) = \left((\alpha_1^*, \ldots, \alpha_m^*), \gamma^*\right)$,is a solution to the dual problem. If there exists a component of α^*, $\alpha_j^* \in \ (0, C)$, then a solution (w^*, b^*) to the problem can be obtained by:

$$w^* = \frac{\gamma^*}{\left(\gamma^* - \sum_{i=1}^{m} \delta_i \alpha_i^*\right)} \sum_{i=1}^{m} \alpha_i^* y_i x_i \,,$$

$$b^* = y_j - \frac{\gamma^*}{\left(\gamma^* - \sum_{i=1}^{m} \delta_i \alpha_i^*\right)} \sum_{i=1}^{m} \alpha_i^* y_i \left({x_i}^T x_j\right) - y_j \delta_j \gamma^*$$

Robust SVC with Ellipsoidal Uncertainty Case Nonlinear

It is easy to extend the problem to the nonlinear case.

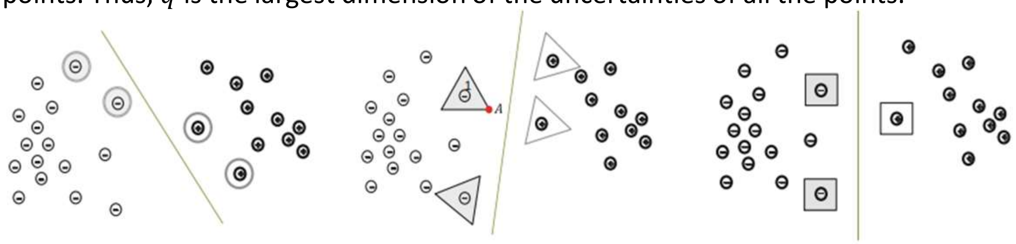
We only need to introduce the kernel function $K\,(x, x')$, and change the inner products (x_i, x_j) to $K\,(x_i, x_j) = \varphi(x_i)^T \varphi(x_j)$

Robust SVC with Polyhedral Uncertainty

In the following, we use a set of inequalities to denote the polyhedral uncertainty for x_i. The data point x_i is said to be polyhedral uncertain if it satisfies that:

$$D_i x_i \leq d_i$$

where the matrix D_i has dimension $q \times n$ and the vector d_i has length q.
Zero vectors could be added to obtain the same number q of inequalities for all data points. Thus, q is the largest dimension of the uncertainties of all the points.



(a) Ellipsoidal Uncertainty (b) Polyhedral Uncertainty ($q = 3$) (c) Interval Uncertainty ($q = 4$)

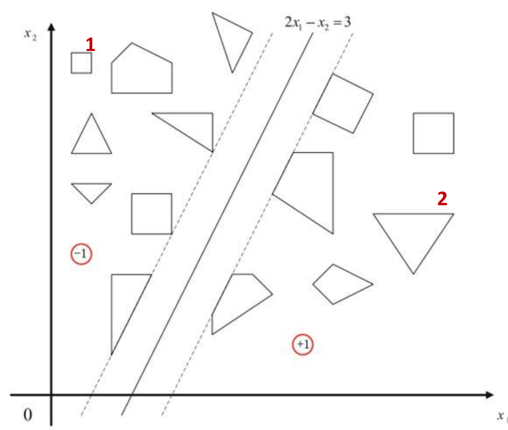
Robust SVC with Polyhedral Uncertainty

Let

$$D_1 = \begin{pmatrix} -1 & 0 \\ 1 & 0 \\ 0 & -1 \\ 0 & 1 \\ 0 & 0 \end{pmatrix},\ d_1 = \begin{pmatrix} 1 \\ -4 \\ 4.25 \\ 0 \\ 0 \end{pmatrix} \ and\ D_2 = \begin{pmatrix} 0 & 1 \\ -3 & -2 \\ 3 & -2 \\ 0 & 0 \\ 0 & 0 \end{pmatrix},\ d_2 = \begin{pmatrix} 2.25 \\ -18 \\ 12 \\ 0 \\ 0 \end{pmatrix}$$

Acording to the definition the two data point are defined as follows:

$$x_1: \begin{cases} -1[x]_1 \le 1 \\ 1[x]_1 \le 4 \\ -1[x]_2 \le 4{,}25 \\ [x]_2 \le 0 \end{cases} \ and\ x_2: \begin{cases} 1[x]_2 \le 1 \\ -3[x]_1 - 2[x]_2 \le -18 \\ 3[x]_1 - 2[x]_2 \le 4{,}25 \end{cases}$$



$(m = 14, n = 2, q = 5)$

Robust SVC with Polyhedral Uncertainty

The robust SVC with polyhedral uncertainty is:

$$\min_{w,b,\xi_i} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m} \xi_i$$

$$s.t. \min_{\{x_i : \boldsymbol{D_i x_i \leq d_i}\}} y_i(w^T x_i + b) \geq 1 - \xi_i,$$

$$\xi_i \geq 0, \qquad i = 1, \ldots, m$$

For example the point with $y_i = -1$, we have that the minimum value $y_i(w^T x_i + b)$ in the region $\{xi : D_i x_i \leq d_i\}$ is larger than 1, or equivalently, the maximum value of $(w^T x_i + b)$ in the region $\{xi : D_i x_i \leq d_i\}$ is less than −1, which is to obtain a separation hyperplane in the worst case for this point.

Robust SVC with Polyhedral Uncertainty

Since $\min_{\{x_i : D_i x_i \le d_i\}} y_i(w^T x_i + b) \ge 1 - \xi_i$ is equivalent to $\max_{\{x_i : D_i x_i \le d_i\}} (-y_i w^T x_i) - y_i\, b \le -1 + \xi_i$

To solve:

$$max\ -y_i\ w^T x_i$$
$$s.t.\ D_i x_i \le d_i$$

The dual is:

$$min\ {d_i}^T z_i$$
$$s.t.\ {D_i}^T z_i = -y_i w$$
$$z_i = (z_{i1}, \ldots, z_{iq})^T \ge 0$$

Robust SVC with Polyhedral Uncertainty

Strong duality would guarantee that the objective values of the dual and primal are equal. Assume that x_i^* and y_i^* are optimal solutions for the previous problems, we have $-y_i w^T x_i^* = {d_i}^T y_i^*$. Therefore, the robust SVM with polyhedral uncertainty formulation is equivalent to:

$$\min_{w,b,\xi_i} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m} \xi_i$$

$$s.t.\ y_i b - {d_i}^T z_i \geq 1 - \xi_i,$$

$${D_i}^T z_i + y_i w = 0, \qquad z_i = (z_{i1}, \dots, z_{iq})^T$$

$$z_{ij} \geq 0, \qquad i = 1, \dots, m, \qquad j = 1, \dots, q$$

$$\xi_i \geq 0, \qquad i = 1, \dots, m$$

Robust SVC with Polyhedral Uncertainty

The Lagrange function for this problem is as follows:

$$L = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m}\xi_i + \sum_{i=1}^{m}\lambda_i\left(y_i b - {d_i}^T z_i - 1 + \xi_i\right) + \sum_{i=1}^{m}\mu_i\left({D_i}^T z_i + y_i w\right) + \sum_{i=1}^{m}\eta_i\xi_i$$

Where $\lambda, \mu, \eta \in \mathbb{R}^m$ are Lagrange multipliers.

Robust SVC with Polyhedral Uncertainty

The duality can be formulated as follow:

$$\max_{\lambda,\mu} \sum_{i=1}^{m} \lambda_i + \frac{1}{2}\sum_{i=1}^{n}\left(\sum_{i=1}^{m} y_i\,\mu_{ik}\right)^2$$

$$s.t.\ y_i d_{i_j} + \sum_{k=1}^{n} \mu_{ik} D_{i_{jk}} = 0, \qquad i = 1,\cdots,m, \qquad j = 1,\cdots,q,$$

$$\sum_{i=1}^{m} \lambda_i y_i = 0,$$

$$0 \le \lambda_i \le C, i = 1,\cdots,m,$$

Robust SVC with Polyhedral Uncertainty

Theorem:

Suppose that $\left(\lambda^{*T}, \mu^{*T}\right) = \left((\lambda_1^*, \ldots, \lambda_m^*), (\mu_1^*, \ldots, \mu_m^*)\right)$,is a solution to the dual problem. If there exists a component of λ^*, $\lambda_j^* \in (0, C)$, then a solution (w^*, b^*) to the problem can be obtained by:

$$w^* = -\sum_{i=1}^{m} \mu_i^* y_i$$

$$b^* = -y_j \sum_{i=1}^{q} (d_i \lambda_i^* + 1)$$

Robust SVC with Polyhedral Uncertainty

The interval uncertainty $[\bar{x}_i - \delta_i, \bar{x}_i + \delta_i]$ is a special case of polyhedral uncertainty since when defining

$$D_i = \begin{pmatrix} I \\ -I \end{pmatrix}, d_i = \begin{pmatrix} \bar{x}_i + \delta_i \\ -\bar{x}_i + \delta_i \end{pmatrix}$$

$\{x_i : x_i \epsilon [\bar{x}_i - \delta_i, \bar{x}_i + \delta_i]\}$ and $\{x_i : D_i x_i \leq d_i\}$ are equivalent.

Uncertainty Support Vector Classification (USVC)

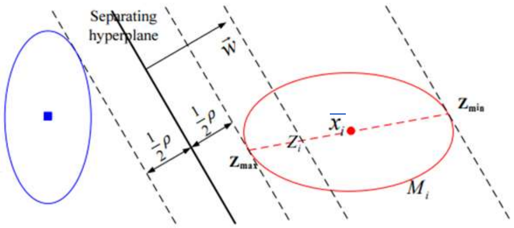
In this case we assume that $x_i \sim \mathcal{N}(\overline{x_i}, M_i)$ in which $\mathcal{N}$ is a Gaussian distribution with mean $\overline{x_i} \in \mathbb{R}^n$ and covariance $M_i \in \mathbb{R}^{n \times n}$:

$$\min_{w,b,\xi_i} \frac{1}{2}\|w\|_2^2 + C\sum_{i=1}^{m} \xi_i$$

$$s.t.\ \ y_i(w^T\overline{x}_i + b) - r\left\|M_i^{1/2}w\right\|_2^2 \geq 1 - \xi_i$$

$$\xi_i \geq 0, \qquad i = 1, \dots, m$$

$$r \geq 0$$



where r needs to be set in advance in the optimization.