Supervised, Unsupervised and Reinforcement Learning in Finance

Week 1: Supervised Learning

Support Vector Machines part II: the math of SVMs

Igor Halperin

NYU Tandon School of Engineering, 2017

Karush-Kuhn-Tucker (KKT) conditions

Our general setting is of a **constrained optimization** of the form

$$\min f(x), \ \ x \in R^n$$
$$s.t. \ \ g(x) \geq 0$$

:

Karush-Kuhn-Tucker (KKT) conditions

Our general setting is of a **constrained optimization** of the form

$$\min f(x),\ \ x \in R^n$$
$$s.t.\ \ g(x) \geq 0$$

Use the method of **Lagrange multipliers** with the **KKT conditions**:

$$L(x,\lambda) = f(x) - \lambda g(x)$$
$$s.t.\ \ g(x) \geq 0, \quad \lambda \geq 0, \quad \lambda g(x) = 0$$

:

Karush-Kuhn-Tucker (KKT) conditions

Our general setting is of a **constrained optimization** of the form

$$\min f(x), \ \ x \in R^n$$
$$s.t. \ \ g(x) \geq 0$$

Use the method of **Lagrange multipliers** with the **KKT conditions**:

$$L(x,\lambda) = f(x) - \lambda g(x)$$
$$s.t. \ \ g(x) \geq 0, \quad \lambda \geq 0, \quad \lambda g(x) = 0$$

Two scenarios for an optimal point x^*:

1) **Inactive constraint**: $g(x) \geq 0, \quad \lambda = 0, \quad \lambda g(x) = 0$

2) **Active constraint**: $g(x) = 0, \quad \lambda > 0, \quad \lambda g(x) = 0$

When the constraint is inactive for an optimal solution, it means that the constraint plays no role, and the solution to the constrained problem is the same as for the unconstrained one.

Constrained optimization for SMV

Training of SVM Regression amounts to **constrained optimization**

$$\min_{w,b,\alpha,\eta} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\left(\xi_i + \xi_i^*\right) - \sum_{i=1}^{N}\left(\eta_i\xi_i + \eta_i^*\xi_i^*\right)$$

$$-\sum_{i=1}^{N}\alpha_i\left(\varepsilon + \xi_i - y_i + \langle w,x\rangle + b\right) - \sum_{i=1}^{N}\alpha_i^*\left(\varepsilon + \xi_i^* + y_i - \langle w,x\rangle - b\right)$$

$$s.t.\ \ \alpha_i,\ \alpha_i^* \geq 0, \quad \eta_i,\ \eta_i^* \geq 0$$

Constrained optimization for SVM

Training of SVM Regression amounts to **constrained optimization**

$$\min_{w,b,\alpha,\eta} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\left(\xi_i + \xi_i^*\right) - \sum_{i=1}^{N}\left(\eta_i\xi_i + \eta_i^*\xi_i^*\right)$$

$$-\sum_{i=1}^{N}\alpha_i\left(\varepsilon + \xi_i - y_i + \langle w,x\rangle + b\right) - \sum_{i=1}^{N}\alpha_i^*\left(\varepsilon + \xi_i^* + y_i - \langle w,x\rangle - b\right)$$

$$s.t.\ \ \alpha_i,\ \alpha_i^* \geq 0, \quad \eta_i,\ \eta_i^* \geq 0$$

Dual optimization problem for Lagrange multipliers $\alpha_i,\ \alpha_i^* \geq 0$:

$$\min_{\alpha_i,\alpha_i^*} \frac{1}{2}\sum_{i,j=1}^{N}\left(\alpha_i - \alpha_i^*\right)\left(\alpha_j - \alpha_j^*\right)\langle x_i, x_j\rangle + \varepsilon\sum_{i=1}^{N}\left(\alpha_i + \alpha_i^*\right) - \sum_{i=1}^{N} y_i\left(\alpha_i - \alpha_i^*\right)$$

$$s.t.\ \sum_{i=1}^{N}\left(\alpha_i - \alpha_i^*\right) = 0, \quad 0 \leq \alpha_i,\ \alpha_i^* \leq C$$

Constrained optimization for SVM

Training of SVM Regression amounts to **constrained optimization**

$$\min_{w,b,\alpha,\eta} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\left(\xi_i + \xi_i^*\right) - \sum_{i=1}^{N}\left(\eta_i\xi_i + \eta_i^*\xi_i^*\right)$$

$$-\sum_{i=1}^{N}\alpha_i\left(\varepsilon + \xi_i - y_i + \langle w,x\rangle + b\right) - \sum_{i=1}^{N}\alpha_i^*\left(\varepsilon + \xi_i^* + y_i - \langle w,x\rangle - b\right)$$

$$s.t.\ \ \alpha_i,\ \alpha_i^* \geq 0, \quad \eta_i,\ \eta_i^* \geq 0$$

Dual optimization problem for Lagrange multipliers $\alpha_i,\ \alpha_i^* \geq 0$:

$$\min_{\alpha_i,\alpha_i^*} \frac{1}{2}\sum_{i,j=1}^{N}\left(\alpha_i - \alpha_i^*\right)\left(\alpha_j - \alpha_j^*\right)\langle x_i, x_j\rangle + \varepsilon\sum_{i=1}^{N}\left(\alpha_i + \alpha_i^*\right) - \sum_{i=1}^{N} y_i\left(\alpha_i - \alpha_i^*\right)$$

$$s.t.\ \sum_{i=1}^{N}\left(\alpha_i - \alpha_i^*\right) = 0, \quad 0 \leq \alpha_i,\ \alpha_i^* \leq C$$

The solution is (see A.Smola and B.Scholkopf, “A tutorial on Support Vector Regression” on how to find b using the KKT conditions):

$$f(x) = \sum_{i=1}^{N}\left(\alpha_i - \alpha_i^*\right)\langle x_i, x\rangle + b$$

Support vector expansion

Constrained optimization for SVM

Training of SVM Regression amounts to **constrained optimization**

$$\min_{w,b,\alpha,\eta} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\left(\xi_i+\xi_i^*\right) - \sum_{i=1}^{N}\left(\eta_i\xi_i+\eta_i^*\xi_i^*\right)$$

$$-\sum_{i=1}^{N}\alpha_i\left(\varepsilon+\xi_i-y_i+\langle w,x\rangle+b\right)-\sum_{i=1}^{N}\alpha_i^*\left(\varepsilon+\xi_i^*+y_i-\langle w,x\rangle-b\right)$$

$$s.t.\ \ \alpha_i,\ \alpha_i^*\geq 0,\quad \eta_i,\ \eta_i^*\geq 0$$

Dual optimization problem for Lagrange multipliers $\alpha_i,\ \alpha_i^*\geq 0$:

$$\min_{\alpha_i,\alpha_i^*} \frac{1}{2}\sum_{i,j=1}^{N}\left(\alpha_i-\alpha_i^*\right)\left(\alpha_j-\alpha_j^*\right)\langle x_i,x_j\rangle+\varepsilon\sum_{i=1}^{N}\left(\alpha_i+\alpha_i^*\right)-\sum_{i=1}^{N}y_i\left(\alpha_i-\alpha_i^*\right)$$

$$s.t.\ \sum_{i=1}^{N}\left(\alpha_i-\alpha_i^*\right)=0,\quad 0\leq\alpha_i,\ \alpha_i^*\leq C$$

The solution is (see A.Smola and B.Scholkopf, “A tutorial on Support Vector Regression” on how to find b using the KKT conditions):

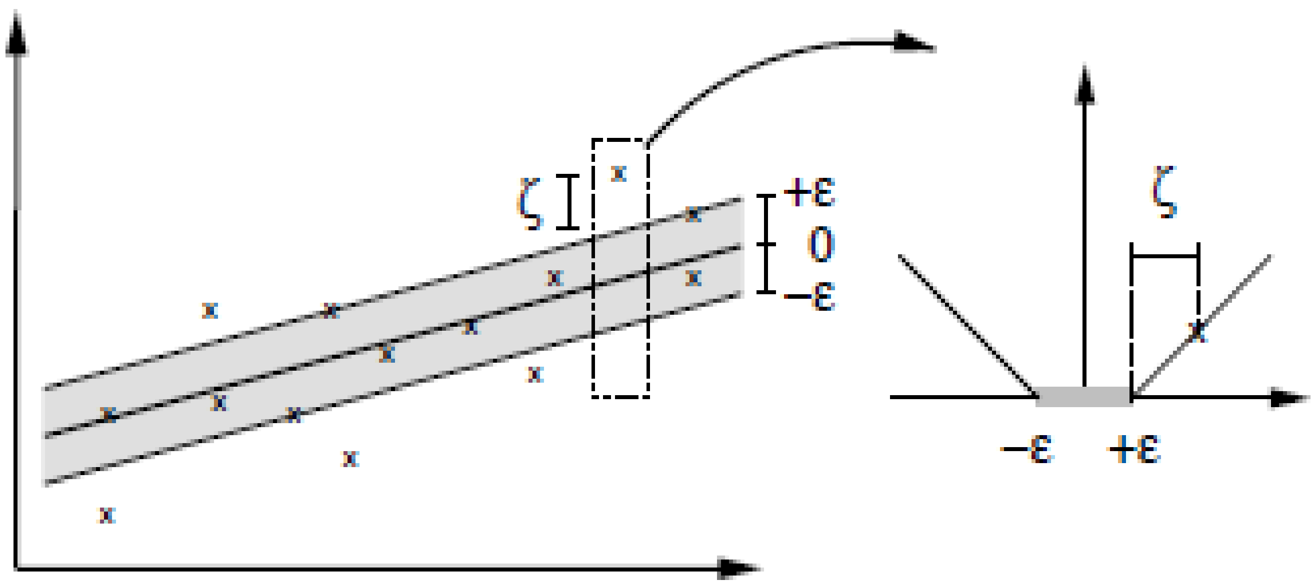
Points within the ε-tube drop from the sum!

$$f(x)=\sum_{i=1}^{N}\left(\alpha_i-\alpha_i^*\right)\langle x_i,x\rangle+b$$

Support vector expansion

Support vector expansion for SVM

- Geometric interpretation: ε-insensitive loss function. Deviations exceeding ε are penalized in a linear fashion



- The solution is found in terms of dual variables α_i, α_i^* (Lagrange multipliers) conjugate to the constraints:

$$f(x) = \sum_i (\alpha_i - \alpha_i^*) \langle x_i, x \rangle + b$$

Inside the ε-tube, α_i, α_i^* vanish (KKT condition), so this is a sparse expansion in *support vectors* - hence the name SVM

Control question

Select all correct answers

1. When a constraint $g(x) \geq 0$ is active for an optimal solution x^*, it means for that solution, we have $g(x^*) \geq 0,\ \lambda = 0$
2. When a constraint $g(x) \geq 0$ is active for an optimal solution x^*, it means for that solution, we have $g(x^*) = 0,\ \lambda > 0$
3. Optimization in SVM amounts to a convex optimization in Lagrange multipliers of the original optimization problem.
4. As training of SVM amounts to convex optimization, it has the same problems with local minima as Neural Networks.
5. Support Vectors are data points outside of the ε-tube.
6. Coefficients of the Support Vector Expansion vanish inside of the ε-tube due to the KKT condition, therefore the Support Vector Expansion is sparse. The complexity of representation in SVM is determined by the number of Support Vectors, rather than by dimensionality of the inputs.

Correct answers: 2, 3, 5, 6.