

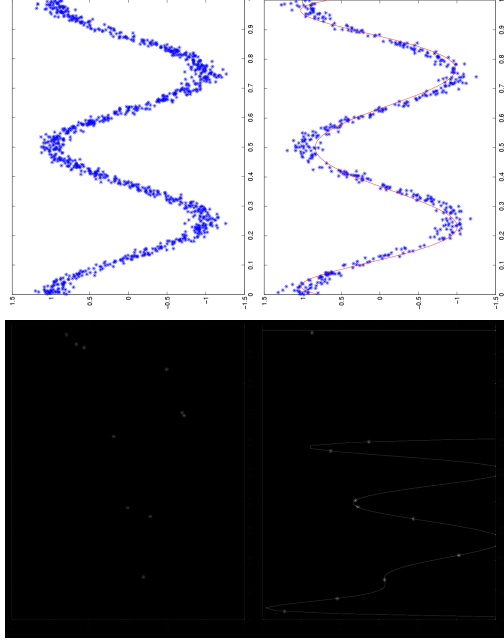
Regression

Data

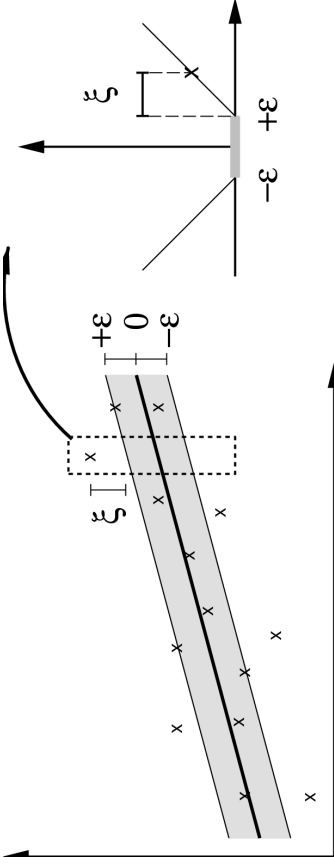
Pairs of observations (x_i, y_i) generated from some $P(x, y)$, e.g., (market index, SP100) (fab parfameters, yield) (user profile, price)

Task

Predict y , given x .



ϵ -insensitive Linear Regressor



Optimization Problem

Find the “flattest” function $f(\mathbf{x}) = \langle \mathbf{x}, \mathbf{w} \rangle + b$ while keeping the approximation error exceeding ϵ , i.e. $|y_i - f(\mathbf{x}_i)|_\epsilon$ as small as possible. Here

$$|\xi|_\epsilon = \max(0, |\xi| - \epsilon) = \begin{cases} |\xi| - \epsilon & \text{if } |\xi| \geq \epsilon \\ 0 & \text{otherwise} \end{cases}$$

Optimization Problem

Idea

We have to rewrite the loss function $|\xi|_\epsilon$ as an optimization problem (week 3).

Analog to Soft Margin Classification

$$\begin{aligned} & \text{minimize} && \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^m (\xi_i + \xi_i^*) \\ & \text{subject to} && ((\mathbf{w}, \mathbf{x}_i) + b) \geq y_i - \epsilon - \xi_i \\ & && ((\mathbf{w}, \mathbf{x}_i) + b) \leq y_i + \epsilon + \xi_i \\ & && \xi_i, \xi_i^* \geq 0 \text{ for all } 1 \leq i \leq m \end{aligned}$$

Interpretation and Regularized Risk Functional

With the loss function $c(\mathbf{x}, y, f(\mathbf{x})) := |y - f(\mathbf{x})|_\epsilon$ this is equivalent to minimizing

$$R_{\text{reg}}[f] = \frac{1}{m} \sum_{i=1}^m |y_i - f(\mathbf{x}_i)|_\epsilon + \frac{\lambda}{2} \|\mathbf{w}\|^2$$

All we have to do is rescale λ into $C = \frac{1}{\lambda m}$.

Lagrange Function and Constraints

Lagrange Function

We have constraints in ξ_i and ξ_i^* , i.e. from both sides, with corresponding η_i, η_i^* .

$$\begin{aligned} L(\mathbf{w}, b, \xi, \xi^*, \alpha, \alpha^*, \eta, \eta^*) &= \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^m (\xi_i + \xi_i^*) - \sum_{i=1}^m (\eta_i \xi_i + \eta_i^* \xi_i^*) \\ &+ \sum_{i=1}^m \alpha_i^* ((\mathbf{w}, \mathbf{x}_i) + b) - y_i - \epsilon - \xi_i^* \\ &+ \sum_{i=1}^m \alpha_i (y_i - \epsilon - \xi_i - ((\mathbf{w}, \mathbf{x}_i) + b)) \end{aligned}$$

Saddlepoint in \mathbf{w}

$$\partial_{\mathbf{w}} L(\mathbf{w}, b, \xi, \xi^*, \alpha, \alpha^*, \eta, \eta^*) = \mathbf{w} + \sum_{i=1}^m (\alpha_i^* \mathbf{x}_i - \alpha_i \mathbf{x}_i) = 0 \iff \mathbf{w} = \sum_{i=1}^m (\alpha_i - \alpha_i^*) \mathbf{x}_i$$

Lagrange Function and Constraints

Saddlepoint in b

$$\partial_b L(\mathbf{w}, b, \xi, \xi^*, \alpha, \alpha^*, \eta, \eta^*) = \sum_{i=1}^m \alpha_i^* - \alpha_i = 0$$

Saddlepoint in ξ_i

$$\partial_{\xi_i} L(\mathbf{w}, b, \xi, \xi^*, \alpha, \alpha^*, \eta, \eta^*) = C - \eta_i - \alpha_i = 0$$

Saddlepoint in ξ_i

$$\partial_{\xi_i^*} L(\mathbf{w}, b, \xi, \xi^*, \alpha, \alpha^*, \eta, \eta^*) = C - \eta_i^* - \alpha_i^* = 0$$

Strategy

Substitute the equations into L to get rid of all primal variables.

Interpretation

Solution in \mathbf{w}

- \mathbf{w} is given by a linear combination of training patterns \mathbf{x}_i and the solution is **independent of the dimensionality** of \mathcal{X} .
- The expansion of \mathbf{w} depends on the Lagrange multipliers α_i and α_i^* .

Kuhn-Tucker-Conditions

We know that at the optimal solution

$$\text{Constraint} \cdot \text{Lagrange Multiplier} = 0$$

Only points with $|y_i - f(\mathbf{x}_i)| \geq \varepsilon$ contribute to the solution, since

$$\alpha_i(y_i - \varepsilon - \xi_i - (\langle \mathbf{w}, \mathbf{x}_i \rangle + b)) = 0 \text{ and } \alpha_i^*(y_i - \varepsilon - \xi_i^* - b) = 0$$

Moreover, $\alpha_i = C$ (and likewise α_i^*) only if $|y_i - f(\mathbf{x}_i)| > \varepsilon$, since also

$$\eta_i \xi_i = (C - \alpha_i) = 0 \text{ and } \eta_i^* \xi_i^* = (C - \alpha_i^*) = 0$$

Only \mathbf{x}_i at or beyond the decision boundary can contribute to \mathbf{w} .

This also allows us to compute b via $b = y_i - \varepsilon - \langle \mathbf{w}, \mathbf{x}_i \rangle$ for $\alpha_i \in (0, C)$.

Dual Optimization Problem

Rewriting the Lagrange Function

$$L = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^m (\alpha_i - \alpha_i^*) \langle \mathbf{x}_i, \mathbf{w} \rangle + \sum_{i=1}^m (\alpha_i - \alpha_i^*) y_i - \sum_{i=1}^m (\alpha_i + \alpha_i^*) \varepsilon$$

$$\sum_{i=1}^m [\xi_i (C - \eta_i - \alpha_i) + \xi_i^* (C - \eta_i^* - \alpha_i^*)] + b \sum_{i=1}^m (\alpha_i^* - \alpha_i)$$

Dual Objective Function

$$D = -\frac{1}{2} \sum_{i,j=1}^m (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) \langle \mathbf{x}_i, \mathbf{x}_j \rangle + \sum_{i=1}^m (\alpha_i - \alpha_i^*) y_i - \sum_{i=1}^m (\alpha_i + \alpha_i^*) \varepsilon$$

Dual Constraints $\mathbf{w} = \sum_{i=1}^m (\alpha_i - \alpha_i^*) \mathbf{x}_i$ and $\sum_{i=1}^m (\alpha_i - \alpha_i^*) = 0$

From $\alpha_i, \eta_i \geq 0$ and $C = \alpha_i + \eta_i$ we conclude $\alpha_i \in [0, C]$.

Kernels

Nonlinearity via Feature Maps

In the linear optimization problem

$$\text{minimize } \frac{1}{2} \sum_{i,j=1}^m (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) \langle \mathbf{x}_i, \mathbf{x}_j \rangle - \sum_{i=1}^m (\alpha_i - \alpha_i^*) y_i + \sum_{i=1}^m (\alpha_i + \alpha_i^*) \varepsilon$$

subject to $\sum_{i=1}^m (\alpha_i - \alpha_i^*) = 0$ and $\alpha_i, \alpha_i^* \in [0, C]$ for all $1 \leq i \leq m$

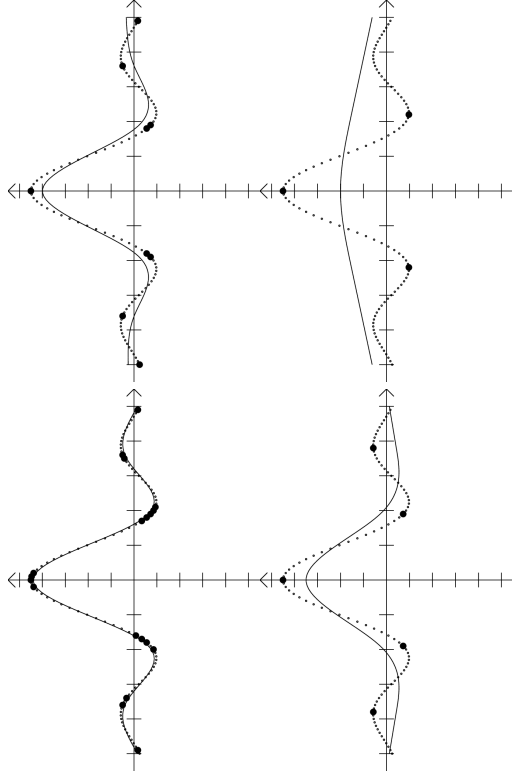
we replace \mathbf{x}_i by $\Phi(\mathbf{x}_i)$ to obtain the new objective function

$$\text{minimize } \frac{1}{2} \sum_{i,j=1}^m (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) k(\mathbf{x}_i, \mathbf{x}_j) - \sum_{i=1}^m (\alpha_i - \alpha_i^*) y_i + \sum_{i=1}^m (\alpha_i + \alpha_i^*) \varepsilon$$

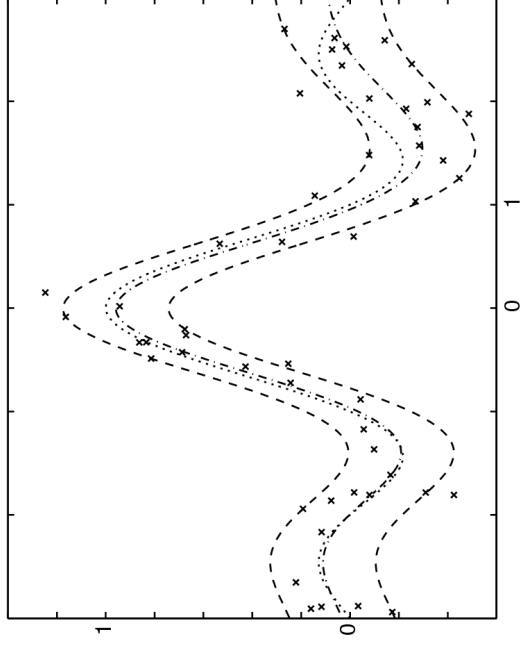
Function Expansion

$$\mathbf{w} = \sum_{i=1}^m (\alpha_i - \alpha_i^*) \Phi(\mathbf{x}_i) \implies f(\mathbf{x}) = \langle \mathbf{w}, \Phi(\mathbf{x}) \rangle + b = \sum_{i=1}^m (\alpha_i - \alpha_i^*) k(\mathbf{x}_i, \mathbf{x}) + b.$$

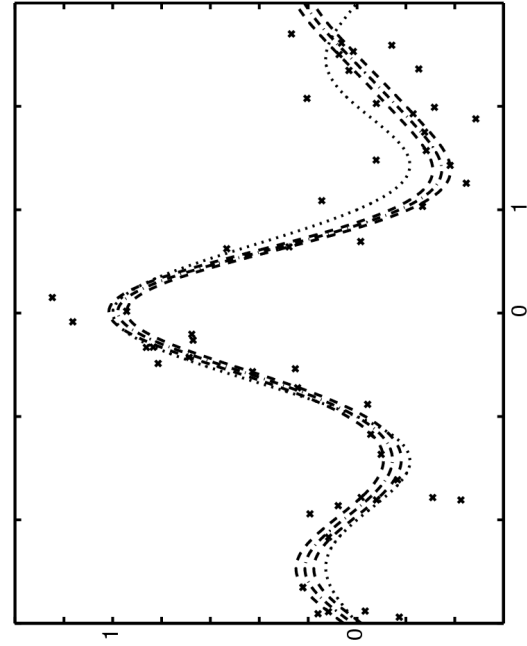
Examples



Examples



Examples



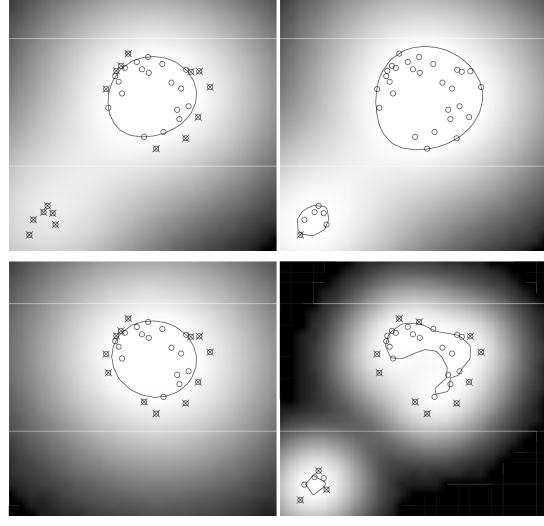
Novelty Detection

Data

Observations (x_i, y_i) generated from some $P(x)$, e.g.,
 (network usage patterns)
 (handwritten digits)
 (alarm sensors)
 (factory status)

Task

Find unusual events, clean database, distinguish typical examples.



Maximum Distance Hyperplane

Idea

Find hyperplane that has **maximum distance from origin** yet is still closer to the origin than the observations.



Hard Margin

$$\begin{aligned} & \text{minimize} && \frac{1}{2} \|\mathbf{w}\|^2 \\ & \text{subject to} && \langle \mathbf{w}, \mathbf{x}_i \rangle \geq 1 \end{aligned}$$

Soft Margin

$$\begin{aligned} & \text{minimize} && \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^m \xi_i \\ & \text{subject to} && \langle \mathbf{w}, \mathbf{x}_i \rangle \geq 1 - \xi_i \\ & && \xi_i \geq 0 \end{aligned}$$

Lagrange Function

Primal Problem

$$\begin{aligned} & \text{minimize} && \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^m \xi_i \\ & \text{subject to} && \langle \mathbf{w}, \mathbf{x}_i \rangle - 1 + \xi_i \geq 0 \text{ and } \xi_i \geq 0 \end{aligned}$$

Lagrange Function

As before, we add the negative constraints to the objective function and obtain:

$$L(\mathbf{w}, \xi) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^m \xi_i - \sum_{i=1}^m \alpha_i (\langle \mathbf{w}, \mathbf{x}_i \rangle - 1 + \xi_i) - \sum_{i=1}^m \eta_i \xi_i \text{ where } \alpha_i, \eta_i \geq 0$$

For optimality we have to compute the partial derivatives of L with respect to \mathbf{w} and ξ and eliminate the primal variables.

Note that we have no constant offset b here.

The Dual Optimization Problem

Optimality Conditions

$$\begin{aligned} \partial_{\mathbf{w}} L = \mathbf{w} - \sum_{i=1}^m \alpha_i \mathbf{x}_i = 0 & \implies \mathbf{w} = \sum_{i=1}^m \alpha_i \mathbf{x}_i \\ \partial_{\xi_i} L = C - \alpha_i - \eta_i = 0 & \implies \alpha_i \in [0, C] \end{aligned}$$

Now we **substitute** the two optimality conditions **back into** L .

Dual Problem

$$\begin{aligned} & \text{minimize} && \frac{1}{2} \sum_{i=1}^m \alpha_i \alpha_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle - \sum_{i=1}^m \alpha_i \\ & \text{subject to} && \alpha_i \in [0, C] \end{aligned}$$

With Kernels

$$\begin{aligned} & \text{minimize} && \frac{1}{2} \sum_{i=1}^m \alpha_i \alpha_j k(\mathbf{x}_i, \mathbf{x}_j) - \sum_{i=1}^m \alpha_i \\ & \text{subject to} && \alpha_i \in [0, C] \end{aligned}$$