

Assignment Project Exam Help

Machine learning lecture slides

<https://eduassistpro.github.io>

Add WeChat edu_assist_pr

Classification II: Margins and SVMs

Assignment Project Exam Help

<https://eduassistpro.github.io>

Add WeChat edu_assist_pr

- ▶ Perceptron
- ▶ Margins
- ▶ Support vector machines
- ▶ Soft-margin SVM

Assignment Project Exam Help

<https://eduassistpro.github.io>

Add WeChat edu_assist_pr

Perceptron (1)

- ▶ Perceptron: a variant of SGD

- ▶ Uses hinge loss: $\ell_{\text{hinge}}(s) := \max\{0, 1 - s\}$

- ▶ Uses conservative updates: only update when there is classification mistake

- ▶ Step size $\eta = 1$

<https://eduassistpro.github.io>

Add WeChat edu_assist_pro

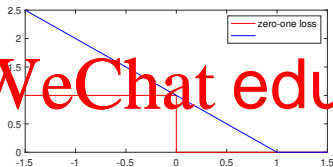


Figure 1: Comparing hinge loss and zero-one loss

Perceptron (2)

- ▶ Start with $w^{(0)} = 0$.
- ▶ For $t = 1, 2, \dots$ until all training examples correctly classified by current linear classifier:
 - ▶ Pick a training example—call it (x_t, y_t) —misclassified by $(t-1)$

Assignment Project Exam Help

<https://eduassistpro.github.io>

Add WeChat edu_assist_pro

Perceptron (3)

- Note that whenever $y_t x_t^\top w^{(t-1)} \leq 0$,

Assignment Project Exam Help

- So update is

- <https://eduassistpro.github.io>

$$\hat{w} = \sum_{i \in S}$$

Add WeChat edu_assist_pr

for some multiset S of $\{1, \dots, n\}$

- Possible to include same example index multiple times in S

Properties of Perceptron

- ▶ Suppose $(x_1, y_1), \dots, (x_n, y_n) \in \mathbb{R}^d \times \{-1, +1\}$ is linearly separable.
- ▶ Does Perceptron find a linear separator? (Yes.) How quickly?
- ▶ Depends on margin achievable on the data set—how much

<https://eduassistpro.github.io>

Add WeChat edu_assist_pro



Figure 2: Linearly separable data

Margins (1)

- Margin achieved by w on i -th training example is the distance from $y_i x_i$ to decision boundary:

$$\gamma_i(w) := \frac{y_i x_i^\top w}{\|w\|_2}.$$

- <https://eduassistpro.github.io>

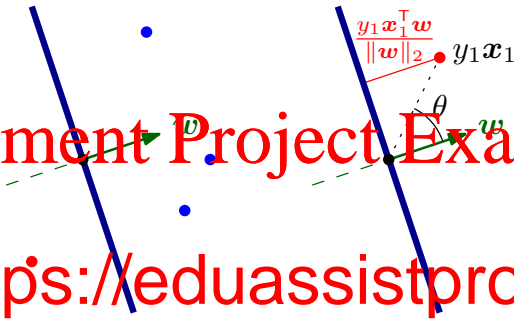
- **Theorem** If training data is linearly separable, then there exists a linear separator after making at most L rescalings, where $L = \max_i \|x_i\|_2$.

Assignment Project Exam Help

<https://eduassistpro.github.io>

Figure 3: Margins

Add WeChat edu_assist_pr



Margins (2)

- ▶ Let w be a linear separator:

$$y_i x_i^\top w > 0, \quad i = 1, \dots, n.$$

- ▶ Note: Scaling of w does not change margin achieved on i th example

- ▶ <https://eduassistpro.github.io>

- ▶ So x_1 is closest to decision boundary among examples.

- ▶ Rescale w so that $y_1 x_1^\top w = 1$
- ▶ Distance from $y_1 x_1$ to decision boundary
- ▶ The shortest w satisfying

$$y_i x_i^\top w \geq 1, \quad i = 1, \dots, n$$

gives the linear separator with the maximum margin on all training examples.

- ▶ Weight vector of maximum margin linear separator: defined as solution to optimization problem

Assignment Project Exam Help

$$\min_{w \in \mathbb{R}^d} \frac{1}{2} \|w\|_2^2$$

<https://eduassistpro.github.io>

- ▶ This is the support vector machine (SVM) problem.
- ▶ Feasible when data are linearly separable
- ▶ Note: Preference for the weight vector achieving maximum margin is another example of inductive bias.

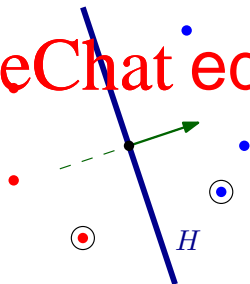
Add WeChat edu_assist_pro

Support vectors

- ▶ Just like least norm solution to normal equations (and ridge regression), solution w to SVM problem can be written as $\sum_{i=1}^n \alpha_i y_i x_i$ for some $\alpha_1, \dots, \alpha_n \in \mathbb{R}$ (in fact, $\alpha_i \geq 0$)
 - ▶ (Adding $r \in \mathbb{R}^d$ orthogonal to span of x_i 's to weight vector can only increase the length without changing the constraint values.)

<https://eduassistpro.github.io>

Add WeChat edu_assist_pr



Soft-margin SVM (1)

- ▶ What if not linearly separable? SVM problem has no solution.
- ▶ Introduce slack variables for constraints, and $C \geq 0$:

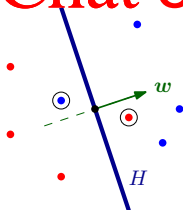
$$\min_{w \in \mathbb{R}^d, \xi_1, \dots, \xi_n \geq 0} \frac{1}{2} \|w\|_2^2 + C \sum_{i=1}^n \xi_i$$

- ▶ <https://eduassistpro.github.io>

- ▶ For given w , $\xi_i / \|w\|_2$ is distance that

$$y_i x_i^T w \geq 1.$$

Add WeChat edu_assist_pr



Soft-margin SVM (2)

- Equivalent unconstrained form:

$$\min_{w \in \mathbb{R}^d} \frac{1}{2} \|w\|_2^2 + C \sum_{i=1}^n \max\{0, 1 - y_i x_i^T w\}$$



<https://eduassistpro.github.io>

$$w \in \mathbb{R}^d \quad \frac{1}{n} \sum_{i=1}^n \max\{0, 1 - y_i x_i^T w\}$$

- Same template as ridge regression Lasso, ...
 - Data fitting term (using a surrogate loss function)
 - Regularizer that promotes inductive bias
 - λ controls trade-off of concerns
- Both SVM and soft-margin SVM can be kernelized