COL774

Machine Learning

Sep 17, 2021

Minor Exam:

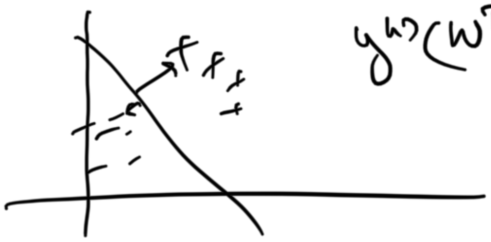
9am - 11am

Tue Sep 21st.

Last class:

SVMs:- Support Vector Machines

$$\left.\begin{array}{l}\min\limits_{w,b} \ \tfrac{1}{2} w^T w \\ y^{(i)}(w^T x^{(i)} + b) \geq 1\end{array}\right\}$$



$$\left[\begin{array}{l}\hookrightarrow \min\limits_{w} \ f(w) \\ \qquad g_i(w) \leq 0 \quad i \in \{1 \ldots m\} \\ \qquad h_l(w) = 0 \quad l \in \{1 \ldots p\}\end{array}\right.$$

$$\hookrightarrow \quad L(w, \alpha, \beta) = f(w) + \sum_{i=1}^{m} \alpha_i g_i(w) + \sum_{l=1}^{p} \beta_l h_l(w)$$

$$\min_{w} \max_{\substack{\alpha,\beta \\ \alpha \geq 0}} L(w,\alpha,\beta) \Leftarrow \left[\min_{w} \ \theta_P(w)\right. \text{ :- Primal problem.}$$

$$\max_{\substack{\alpha,\beta \\ \alpha \geq 0}} \min_{w} L(w,\alpha,\beta) \Leftarrow \max_{\alpha,\beta} \ \theta_D(\alpha,\beta) \text{ :- Dual problem.}$$

$\alpha \geq 0$; $\{\alpha_i \geq 0 \ \forall i\}$

$$\Rightarrow \theta_P(w) \geq \theta_D(\alpha, \beta)$$

$$\Rightarrow p^* = \min_{w} \ \theta_P(w) \geq \max_{\substack{\alpha,\beta \\ \alpha \geq 0}} \theta_D(\alpha, \beta) = d^*$$

Weak duality

Strong duality

$p^* = d^*$?

SVMs: I

(i) Convex linearly

(ii) Data is separable

KKT conditions ✓

(1) Problem is convex

(2) Slaters conditions are satisfied

$\exists w, \forall i \; g_i(w) < 0 \} \; h_l(w) = 0 \; \forall l$

optimal value of primal variables

$\Rightarrow \exists w^*, \alpha^*, \beta^*$ s.t optimal value of dual variables

$p^* = d^* = L(w^*, \alpha^*, \beta^*)$

$\Updownarrow$

KKT conditions (Karush-Kuhn-Tucker)

(I) Gradients vanish at optimal parameters

(a) $\nabla_w L(w, \alpha, \beta) \Big|_{w^*} = 0$

(b) $\nabla_\beta L(w, \alpha, \beta) \Big|_{\beta^*} = 0$

(II) Primal & Dual feasibility

(a) $g_i(w^*) \leq 0 \; \forall i \quad h_l(w^*) = 0 \; \forall l \}$ primal feasibility

(b) $\alpha^* \geq 0 \quad \} \quad \alpha_i^* \geq 0 \; \forall i$

(III) Complementary slackness:-

At w^*:- $\quad \alpha_i \cdot g_i(w) = 0 \quad \forall i$

if $\alpha_i > 0 \Rightarrow g_i(w) = 0 \;]$ ith constraint is tight

And $g_i(w) > 0 \Rightarrow \alpha_i = 0 \;]$ ith constraint is not tight

$\alpha_i > 0$ if $g_i(w) = 0 \geq$

$\Rightarrow y^{(i)}(w^T x^{(i)} + b) = 1 \Rightarrow$ does not play

0

$\hookrightarrow$ ith point lies on margin boundary

a deciding role in determining the value of optimal params

Support Vec

$g_i(w) \geq 0 \Rightarrow$ point is away from the margin boundary.

$\Rightarrow \alpha_i = 0$

SVMs :-

:- strong duality holds

$$L(w, b, \alpha) = \frac{1}{2} w^T w + \underbrace{\sum_{i=1}^{m} \alpha_i \left(1 - y^{(i)}(w^T x^{(i)} + b)\right)}$$

$\hookrightarrow$ Try to get rid of primal variable (get the SVM dual)

$$\max_{\alpha : \alpha \geq 0} \left[\underbrace{\min_{w, b} L(w, b, \alpha)} \right] \quad : \text{SVM dual}$$

$$\nabla_w L(w, b, \alpha) = \frac{1}{\cancel{2}} \cancel{2} w + \sum_{i=1}^{m} \alpha_i (-1) y^{(i)} x^{(i)}$$

Equate $\nabla_w L(w, b, \alpha) = 0$

$$① \quad w = \sum_{i=1}^{m} \alpha_i y^{(i)} x^{(i)}$$

only those constraints play a role which are active: $(g_i(w, b) = 0)$

$$L(w, b, \alpha) = \frac{1}{2} w^T w + \sum_{i=1}^{m} \alpha_i \left(1 - y^{(i)}(w^T x^{(i)} + b)\right)$$

$$\nabla_b L(w, b, \alpha) = 0 + \sum_{i=1}^{m} -\alpha_i y^{(i)}$$

Equate ~~to~~ $\nabla_b L(w, b, \alpha) = 0$

$$② \quad -\sum_{i=1}^{m} \alpha_i y^{(i)} = 0$$

substituting ① & ⑪ in $L(w, b, \alpha)$

$$L(w,b,\alpha) = \frac{1}{2} w^T w + \sum_{i=1}^{m} \alpha_i \left(1 - y^{(i)} (w^T x^{(i)} + b)\right)$$

$$= \frac{1}{2} \left[\sum_{i=1}^{m} \alpha_i y^{(i)} x^{(i)}\right]^T \left[\sum_{j=1}^{m} \alpha_j y^{(j)} x^{(j)}\right]$$

$$+ \sum_{i=1}^{m} \alpha_i \left[1 - y^{(i)} \left[\sum_{j=1}^{m} \alpha_j y^{(j)} x^{(j)}\right]^T x^{(i)} + b\right]$$

$$= \frac{1}{2} \sum_{i,j=1}^{m} \alpha_i \alpha_j y^{(i)} y^{(j)} x^{(i)T} x^{(j)}$$

$$+ \sum_{i=1}^{m} \alpha_i - \sum_{i=1}^{m} \alpha_i y^{(i)} \left[\sum_{j=1}^{m} \alpha_j y^{(j)} x^{(j)T}\right] x^{(i)}$$

$$- \underbrace{\cancel{\sum_{i=1}^{m} \alpha_i y^{(i)}} \cdot b}_{0} \qquad (\text{using } ②)$$

$$= \frac{1}{2} \sum_{i,j=1}^{m} \alpha_i \alpha_j y^{(i)} y^{(j)} x^{(i)T} x^{(j)}$$

$$+ \sum_{i=1}^{m} \alpha_i - \sum_{i,j=1}^{m} \alpha_i \alpha_j y^{(i)} y^{(j)} x^{(i)T} x^{(j)} + 0$$

$$= \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{m} \alpha_i \alpha_j y^{(i)} y^{(j)} x^{(i)T} x^{(j)}$$

$$\equiv W(\alpha) \;\Big]$$ Dual objective.

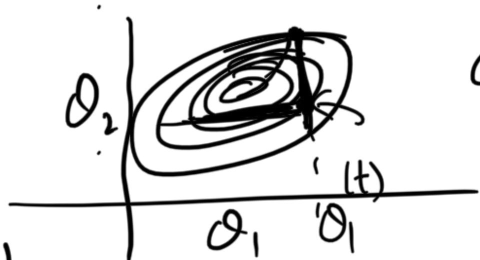
$$\boxed{\begin{array}{l} \text{:- Dual problem:- quadratic} \\ \max_{\alpha} \; W(\alpha) : \\ \alpha_i \geq 0 \;\forall i \quad \sum_{i=1}^{m} \alpha_i y^{(i)} = 0 \end{array}}$$

SVM Dual formulation

↳ Block coordinate Descent

SMO:- Sequential Minimal Optimization

Optimize over two variables at a time.



co-ordinate descent:

→ Picks a coordinate at a time & optimizes wrt.

Quadratic expr. in single variable ⇐

$$\max_{\alpha_1 \ldots \alpha_m} W(\alpha) :-$$

$$\alpha_i \geq 0 \quad \sum_{i=1}^{m} \alpha_i y^{(i)} = 0$$

Fix:- α's except one α_i'

SVMs:- ~~Dual~~ Block co-ordinate Descent

↳ optimize 2 variables at a time.

Repeat this process (α_i, α_j :- Fix all other variables

This until converge for another set of (α_i, α_j) ↓ optimize wrt α_i, α_j

[~~Book~~ Read from Andrew Ng's notes on SVMs]

$$w = \sum_{i=1}^{m} \alpha_i y^{(i)} x^{(i)}$$

$\because \alpha_i > 0$
$\Rightarrow$ i^{th} ~~point~~ constraint is tight
$\Rightarrow$ $x^{(i)}$ is on margin boundary

Support Vectors :- $\alpha_i > 0$

Finally :- How do we find b?

$y^{(i)} = 1$

$$y^{(i)}(w^T x^{(i)} + b) \geq 1 \quad \forall i$$

$$w^T x^{(i)} + b \geq 1$$

$$\left[b \geq 1 - w^T x^{(i)} \right] \quad \forall i :-$$

$$b = 1 - w^T x^{(i)}$$:- for support vectors.

$$b = \max_{i : y^{(i)} = 1} \left[1 - w^T x^{(i)} \right]$$

$$= 1 - \min_{y^{(i)} = 1} w^T x^{(i)} \quad \text{— ⓐ}$$

$y^{(i)} = -1$

$$y^{(i)}(w^T x^{(i)} + b) \geq 1$$

$$(w^T x^{(i)} + b) \le -1$$

$$b \le [-1 - w^T x^{(i)}] \;\forall i$$

$$b = -1 - w^T x^{(i)}$$ for support vectors:-

$$b = \min_{i:y^{(i)}=-1} [-1 - w^T x^{(i)}]$$

$$= \cancel{\min_{\forall i}} -1 - \max_{i:y^{(i)}=-1} w^T x^{(i)}$$

$$\text{—} \quad b$$

$$2b = -\left[\min_{i:y^{(i)}=1} w^T x^{(i)} + \max_{i:y^{(i)}=-1} w^T x^{(i)}\right]$$

$$\boxed{b = -\tfrac{1}{2}\left[\min_{i:y^{(i)}=1} w^T x^{(i)} + \max_{i:y^{(i)}=-1} w^T x^{(i)}\right] \qquad w = \sum_{i=1}^{m} \alpha_i\, y^{(i)} x^{(i)}}$$

SVM for non-linearly sep data

SVM for high dim data