COL774

Machine Learning

Sep 14, 2021

Last class:- Naive Bayes

↳ Text Classification.

$\arg\max_{\theta} P(\theta|D)$:- MAP estimate

as opposed to $P(D;\theta)$ or $P(D|\theta)$

$\arg\max_{\theta} P(D|\theta)$:- ML estimate

Laplace smoothing

Regularization

Imp:- Practical Imp:- Inference time

$\arg\max_{y} P(y|x;\theta)$

$$= \arg\max_{y} \frac{\left[\left[\prod_{j=1}^{n} P(x_j|y;\theta)\right] P(y)\right]}{\cancel{P(x)}}$$

→ Naive Bayes Ass.

$\theta_{j\ell|y=k}$

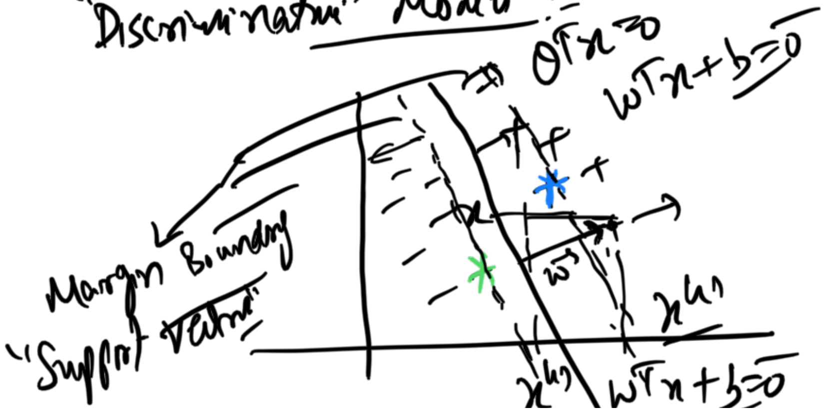
→ underflow

⇒ Do take log before computation.

Today's Class:-

Support Vector Machines (SVMs)

"Discriminative" Models :- Binary Classification.



$y^{(i)} \in \{-1, 1\}$

Linear Separator

$x^{(i)} \in \mathbb{R}^n$ $\quad \theta_0, x_0 = 1$

$\theta^T x = 0$ (included the

Logistic Regression

$P(y^{(i)} | x^{(i)}; \underset{\sim}{\theta})$

$= \frac{1}{1+e^{-\theta^T x^{(i)}}}$:- sigmoid function

$\log \prod_{i=1}^{m} \left[P(y^{(i)} | x^{(i)}; \theta) \right]$

every example contributed to computation of θ parameters.

(intercept term)

$w^T x + b = 0$ → intercept

$w = (\theta_1 \; -- \; \theta_n)$

SVMs :- Support Vectors :- supporting the hyperplane

:- Margin of a point (given $w^T x + b = 0$)

"Signed" Distance of a point from the hyperplane

(normalized distance)

$\gamma^{(i)} \Leftarrow \left[\frac{y^{(i)}(w^T x^{(i)} + b)}{\|w\|_2} \right]$:- unnormalized dist. → normalized Distance

↳ geometric margin

Functional margin $\hat{\gamma}^{(i)}$

:- SVMs find the hyperplane which maximizes the min margin

Max - Margin Based classifiers

↳ $\max_{\gamma, w, b} \gamma$

$\underline{\gamma^{(i)}} \geq \gamma \quad \forall i$

Maximizes the minimum margin

$\gamma^{(i)} = y^{(i)} [(w^T x^{(i)} + b)]$

$$= \left[\frac{\quad}{\|w\|} \right]$$

$$\max_{\gamma, w, b} \quad \gamma$$

$$\frac{y^{(i)}(w^T x^{(i)} + b)}{\|w\|} \geq \gamma$$

$$\hat{\gamma} = \|w\| \gamma$$

$$\max_{\hat{\gamma}, w, b} \quad \frac{\hat{\gamma}}{\|w\|}$$

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

constrained optimization problem

Suppose:- $\hat{\gamma}^*, w^*, b^*$ is an optimal solution to this problem,

$k > 0$ $\quad k\hat{\gamma}^*, kw^*, kb^*$ is also an optimal solution

$$\Rightarrow k = \frac{1}{\hat{\gamma}^*}$$

$$\Rightarrow 1, \frac{w^*}{\hat{\gamma}^*}, \frac{b^*}{\hat{\gamma}^*}$$

$$\Rightarrow 1, w'^*, b'^*$$

$$\Rightarrow \max_{w,b} \quad \frac{1}{\|w\|}$$

$$\text{s.t.} \quad y^{(i)}(w^T x^{(i)} + b) \geq 1$$

equivalent optimization problem

$$\max_{w,b} \frac{1}{\|w\|^2}$$

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

$\|w\|^2 = w^T w$

$$\min_{w,b} \frac{1}{2} w^T w$$

subject to constraint $\forall i$:- $y^{(i)}(w^T x^{(i)} + b) \geq 1$ → $\gamma^{(i)}$: $\underbrace{1 - y^{(i)}(w^T x^{(i)} + b) \leq 0}_{g_i(w,b)}$

Constrained optimization problem

SVM optimization problem

↳ Convex optimization problem (constrained)

$\min_w f(w)$ $w \in \mathbb{R}^n$

$g_i(w) \leq 0$ $\forall i \in \{1 \ldots m\}$ inequality constraint

$h_l(w) = 0$ $\forall l \in \{1 \ldots p\}$ equality constraints

Convex Constrained Optimization :-

if (i) f :- is a convex function

(ii) g_i :- also convex function $[\forall i \in \{1 \ldots m\}]$

(iii) h_l :- are affine :-

$h_l(w) = w^T x + b$

$w^T x + b = 0$ linear

$C\,P\,(\alpha w_1 + \beta w_2)$ ...

↳ convex

↳ concave

$= \alpha h_e(w_1) + \beta h_e(w_2)$ convex

$f(x) = mx + c$

$f(\theta)$:- $f(\alpha\theta_1 + (1-\alpha)\theta_2) \leq \alpha f(\theta_1) + (1-\alpha) f(\theta_2)$: convex fn

↳ Langrangeans: →