Monday, September 29, 2014 3:24 PM

MLE/ MAP [in] variance
Support Vector Machines
Convex functions

"Standardize Features"

- When do MAP, we often standardize columns

$x = \begin{bmatrix} & \bigcirc & \end{bmatrix}$

In regression we sometimes do this to 'Y', so we don't need y-intercept.

subtract mean } mean 0
divide by variance } variance 1

→ roughly puts weights w_i on same side

"Bias" vs "bais"

- sometimes you'll see logistic regression written using

$$p(y_i \mid \bar{x}_i, \bar{w}, b) = \frac{1}{1 + \exp(-y_i(\bar{w}^T\bar{x}_i + b))}$$

← bais variable

instead of

$$p(y_i \mid \bar{x}_i, w) = \frac{1}{1 + \exp(-y_i \bar{w}^T \bar{x}_i)}$$

- "Bais" variable 'b' takes into account that one class may be more likely before we see the features (c.f., $p(y_i = 1)$ in Naive Bayes)
- Equivalent to adding a column of ones to X
- Statisticians say: "do not regularize the bias"

compromize: regularize bais with a smaller amount, larger weight for features

Linear Classifiers

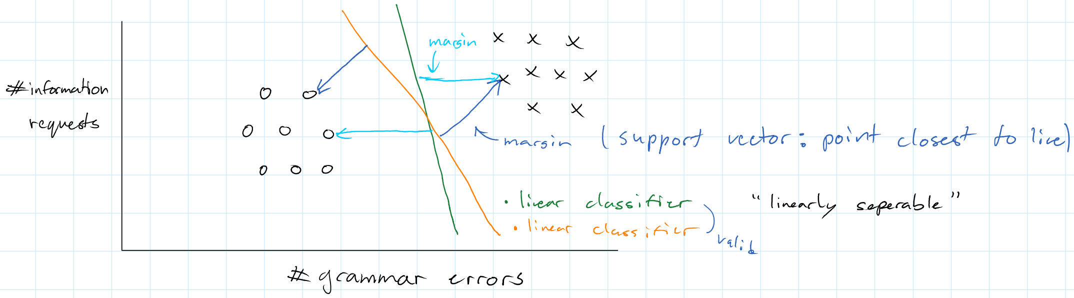
Generative
- LDA
- Naive Bayes

Discriminative
- Logistic Regression

Discriminative Function
- Support vector machines

* today

Support Vector Machines



line is uniquely defined by support vectors (we only need the closest points on each side, can throw away the rest of the data)

Maximum Margin

- maximize distance of line to closest point

To find the optimal line:

"margin" is proportional to $\frac{1}{\|\bar{w}\|}$

(geometric argument, see wikipedia)

maximize $\frac{1}{\|\bar{w}\|} \iff$ minimizing $\|\bar{w}\|^2$

SVM:

$$\min_{w} \frac{1}{2}\|\bar{w}\|^2 \quad \text{s.t.} \quad y_i \bar{w}^T \bar{x}_i - 1 \geq 0$$

↳ when this is equal (roots), will be the support vectors

find best line such that:

o	points on first side	$\bar{w}^T \bar{x}_i \geq 1$	$y_i = 1$
x	points on other side	$\bar{w}^T \bar{x}_i \leq -1$	$y_i = -1$

"Soft-margin" SVM

C   1   equivalent

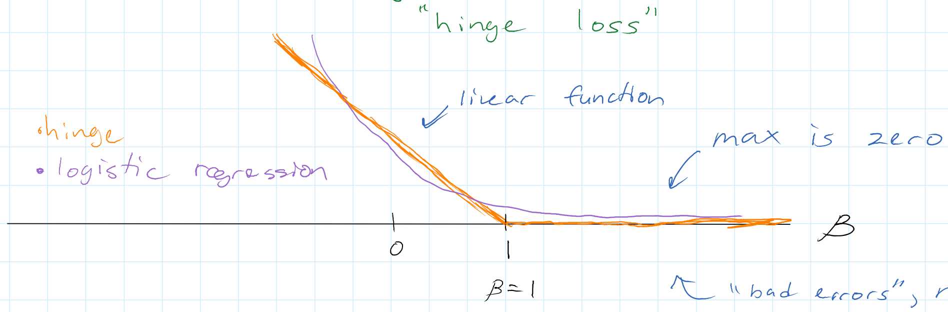
$$\min_{\bar{w}, \bar{V}} \sum_{i=1}^{n} V_i + \frac{\lambda}{2}\|\bar{w}\|^2 \quad \text{s.t.} \quad y_i \bar{w}^T \bar{x}_i - 1 + V_i \geq 0, \quad V_i \geq 0$$

lets you violate the constraint a bit

"for C large enough, linearly seperable is a special case

$$V_i \geq \max\{0, 1 - y_i \bar{w}^T \bar{x}_i\}$$

$$\operatorname*{argmin}_{w} \underbrace{C \sum_{i=1}^{N} \max\{0, 1 - \overbrace{y_i \bar{w}^T \bar{x}_i}^{\beta}\}}_{\text{"hinge loss"}} + \frac{1}{2}\|w\|^2$$



↖ "bad errors", no penalty for being right

Multi-Class SVM

ie $y_i = \{1, 2, 3 \cdots K\}$
or $y_i = \{$red, blue, $\cdots\}$

- 1 vs all: train k binary classifiers $(O(dk))$ 1 vs all, 2 vs all, ...
- all pairs: 1 vs 2
  1 vs 3 ... (doesn't scale well with k)
  2 vs 3 ...
- ECOC error correcting output code

– Multiclass:

$$\min_{w_1, w_2, \ldots w_K} \frac{1}{2}\|\bar{w}\|^2 + C\sum_{i=1}^{N} \max_c \{1 - \bar{w}_{y_i}^T\bar{x}_i + \bar{w}_c^T\bar{x}_i\}$$

$$\sum_{i=1}^{N}\sum_{c=1}^{K} \max\{0, 1 - \bar{w}_{y_i}^T\bar{x}_i + \bar{w}_c^T\bar{x}_i\}$$

Other types of regression / classifications

- Extreme value regression (very rare class, treat asymetrically)
- Support Vector regression

- ordinal logistic regression, proportional hazards (movie ratings)
- Ranking ($a \geq b$)
- Multiple regression
- Multi-task classification (predict multiple y's)

- "Structured prediction"

Convex Functions

one minimal value (maybe a set of locations)

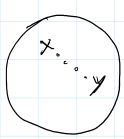
Why? - polynomial-time algorithms to optimize
- stationary points are global optima

$d/dx = 0$

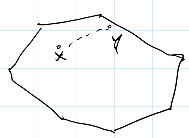
(least squares, L2-reg, L1-reg, logistic, SVMs)

Convex Set

A set C is convex if $\theta x + (1-\theta)y \in C_y$



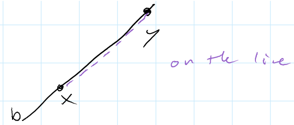
$\forall x, y \in C_y \quad 0 \leq \theta \leq 1$



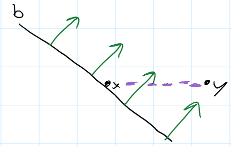
not convex

ie, no bending inwards.

Linear Equality $\{x \mid a^T x = b\}$



Inequality $\{x \mid a^T x \leq b\}$



norm-ball $\{x \mid \|x\| \leq r\}$

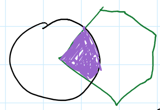
norm-cone $\{(x, r) \mid \|x\| \leq r\}$

Spectrahedron $\{\Sigma \mid \Sigma \succcurlyeq 0\}$
$\{\Sigma \mid \Sigma \succ 0\}$

$\{x \mid f(x) \leq 0\}$

for any convex f.

Intersection of convex sets is convex

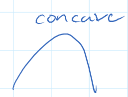


LP $\{x \mid Ax \leq b,\ Aeq = beq,\ LB \leq x \leq UB\}$

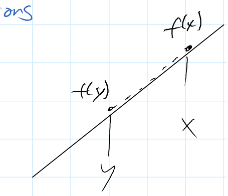
"closed" convex set if it includes boundary

Convex Functions

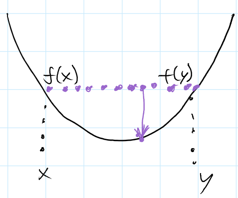
A function 'f' is convex if dom(f) is a convex set, and



linear functions both convex and concave



$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta) f(y)$$



$\forall x \in \mathrm{dom}(f),\ y \in \mathrm{dom}(f),\ 0 \leq \theta \leq 1$

epigraph is a convex set

Examples:

$f(x) = ax^2 + b, \quad a \geq 0$

$f(x) = \exp(ax)$

$f(x) = x \log(x), \quad x \geq 0$