

IFT6132

TODO: fill survey <http://bit.ly/IFT6132-W24> by tomorrow!

10h22

structured prediction basics + setup

Learning problem

given a training dataset $D = (x^{(i)}, y^{(i)})_{i=1}^n$

goal: learn a prediction mapping $h_w: X \rightarrow Y$
 parameter

that has low classification error
generalization error

$$L(w; P) \triangleq \mathbb{E}_{(x,y) \sim P} [\ell(y, h_w(x))]$$

test distribution

structured error function

statistical decision loss

[see lecture 5 of my PGM → review of statistical decision theory]

"risk" in ML (Vapnik)

examples of learning approaches:

regularized ERM
 ↳ ('empirical risk minimization')

$$\hat{L}(w) = \underbrace{\frac{1}{n} \sum_{i=1}^n \ell(y^{(i)}, h_w(x^{(i)}))}_{\text{empirical error}} + \underbrace{R(w)}_{\text{regularizer}}$$

not in ds. in w.

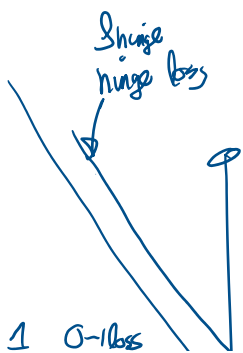
non-convex

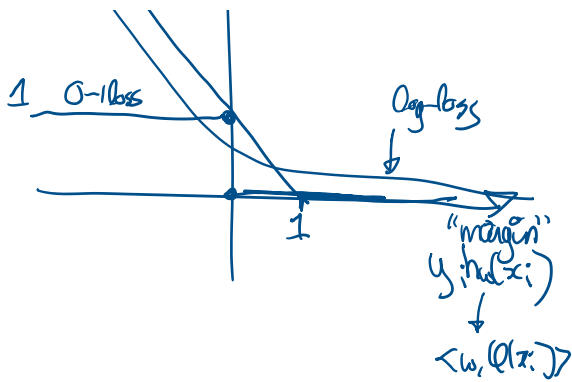
messy [NP hard to minimize in general]

replace \hat{L} with a surrogate loss / contrast f. in statistics

$$\hat{L}(w) = \frac{1}{n} \sum_{i=1}^n \underbrace{\ell(x^{(i)}, y^{(i)}, w)}_{\text{surrogate loss e.g. convex in } w} + R(w)$$

↳ example: • structured loss





Examples: • structured hinge loss
 → structured SVM
 • log-loss → CRF
 aka "cross-entropy"

M-estimator
 in statistics

generative vs discriminative learning continuum

input pair
 (x, y)

generative modeling

$p_w(x, y)$ → conditional $p(y|x)$ → prediction function
 "more discriminative"

model: $p_w(x, y)$ $p_w(y|x)$ $h_w: X \rightarrow \mathcal{Y}$

learning: learn \hat{w}
 by ML

learn \hat{w}
 by MLE

use $l(y, y')$ to construct
 surrogate loss

learn \hat{w} by surrogate loss
 minimization

MRF

CRF

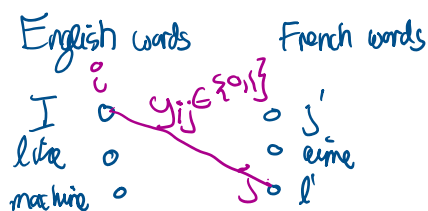
e.g. structured SVM

← more assumptions / less robust on classification tasks

some important aspects of structured prediction (vs binary classification)

- 1) \mathcal{Y} output is usually exponentially big (in natural size of input)
- 2) structured error function $l(y, y')$
- 3) sometimes constraints on pieces of y

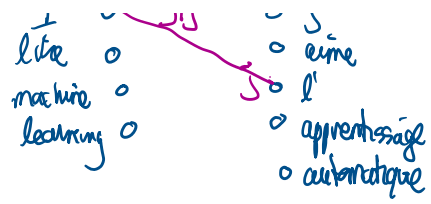
↳ consider word alignment example



note: need "encoding $f(\cdot)$ "

$$y = (y_1, \dots, y_p) \in \mathbb{R}^p$$

$$\rightsquigarrow y = (y_{ij})_{i,j=1}^{|E|} \in \{0,1\}^{|E|}$$



$$\rightsquigarrow y = (y_{ij})_{(i,j) \in E} \in \{0,1\}^{|E|}$$

↑
possible edges (have all pairs)

matching constraints $\left\{ \begin{array}{l} \sum_j y_{ij} \leq 1 \quad \forall i \\ \sum_i y_{ij} \leq 1 \quad \forall j \end{array} \right.$