

- TODO: fill survey <http://bit.ly/IFT6132-W20> by Wednesday evening!

structured prediction basics & setup

Learning problem:

given a training dataset $D = (x^{(i)}, y^{(i)})_{i=1}^n$
 $\downarrow \quad \downarrow$
 $\in X \quad \in Y$

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

that has low classification error
generalization error

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

parameter

test distribution

structured error function

statistical decision loss / "risk" in ML (Vapnik)

[see lecture 4 & 5 of my PGM \rightarrow review of statistical decision theory]

example of learning approach:

regularized ERM
 'empirical risk minimization'

$\hat{L}(w) = \frac{1}{n} \sum_{i=1}^n l(y^{(i)}, h_w(x^{(i)})) + R(w)$

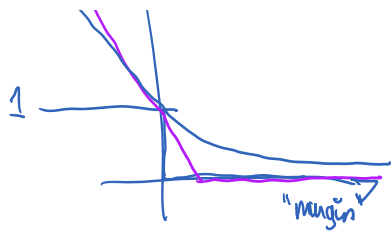
regularizer

not cts in w ; non-convex
 messy [NP hard to minimize in general]



\Rightarrow replace \hat{L} with a surrogate loss / contrast fct:
 ML statistics

$\hat{Q}(w) = \frac{1}{n} \sum_{i=1}^n Q(x^{(i)}, y^{(i)}, w) + R(w)$



$$\mathcal{L}(w) = \frac{1}{n} \sum_{i=1}^n \ell(x^{(i)}, y^{(i)}, w) + R(w)$$

surrogate loss e.g. convex in w

- ↳ examples:
- structured hinge loss → structured SVM
 - log-loss → CRF

M-estimator in statistics

15h31

generative vs. discriminative learning continuum

generative modeling (x, y) ^{input} ^{output}

$p_w(x, y)$ → condition $p(y|x)$ → prediction function

"more discriminative"

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

learning: learn \hat{w} by ML

learn \hat{w} by MCL

use $\ell(y, y')$ to construct surrogate loss
learn \hat{w} by surrogate loss minimization
e.g. structured SVM

more assumptions / less robust on classification task

Some important aspects of structured prediction (vs. binary classification)

1) Y output space is usually exponentially big (in natural size of input)

2) structured error function $\ell(y, y')$

3) sometimes constraints on pieces of y

note: need an "encoding function" $y = (y_1, \dots, y_n) \in \mathbb{R}^n$

↳ consider word alignment example:

English words French words

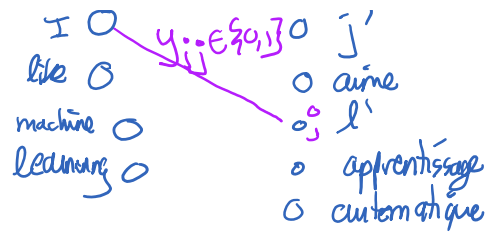
I j

like aime

$y_{ij} \in \{0, 1\}$

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

possible edges (here, all pairs)



possible edges
(here, all pairs)

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