Semi-Supervised learning_1

2021年4月1日 19:43

Soft have no use
Entropy-based regularization

- Supervised learning: $\{(x^r, \hat{y}^r)\}_{r=1}^{R}$
  - E.g. x^r: image, $\hat{y}^r$: class labels
- Semi-supervised learning: $\{(x^r, \hat{y}^r)\}_{r=1}^{R}$, $\{x^u\}_{u=R}^{R+U}$
  - A set of unlabeled data, usually U >> R 没有用testing-set的label所以不是cheating
  - Transductive learning: unlabeled data is the testing data
  - Inductive learning: unlabeled data is not the testing data

Semi-supervised Learning for Generative Model

Low-density Separation Assumption

Smoothness Assumption

Better Representation

The distribution of the unlabeled data tell us ***something***.

Supervised Generative Model

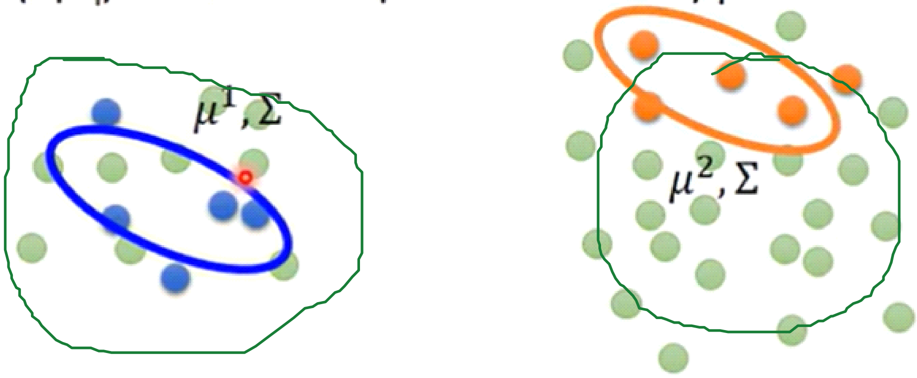
- Given labelled training examples $x^r \in C_1, C_2$
  - looking for most likely prior probability $P(C_i)$ and class-dependent probability $P(x|C_i)$
  - $P(x|C_i)$ is a Gaussian parameterized by μ^i and Σ

With $P(C_1), P(C_2), \mu^1, \mu^2, \Sigma$

$$P(C_1|x) = \frac{P(x|C_1)P(C_1)}{P(x|C_1)P(C_1) + P(x|C_2)P(C_2)}$$

Semi-supervised Generative Model

- Given labelled training examples $x^r \in C_1, C_2$
  - looking for most likely prior probability $P(C_i)$ and class-dependent probability $P(x|C_i)$
  - $P(x|C_i)$ is a Gaussian parameterized by μ^i and Σ



The unlabeled data x^u help re-estimate $P(C_1), P(C_2), \mu^1, \mu^2, \Sigma$

看了unlabel 的 data 会影响对mean 和cov的预测,

进而影响decision boundary

Semi-supervised Generative Model

- Initialization: $\theta = \{P(C_1), P(C_2), \mu^1, \mu^2, \Sigma\}$
- Step 1: compute the posterior probability of unlabeled data

$P_\theta(C_1|x^u)$ Depending on model θ

- Step 2: update model

$$P(C_1) = \frac{N_1 + \boxed{\sum_{x^u} P(C_1|x^u)}}{N}$$

N: total number of examples
N_1: number of examples belonging to C_1

$$\mu^1 = \frac{1}{N_1}\sum_{x^r \in C_1} x^r + \boxed{\frac{1}{\sum_{x^u} P(C_1|x^u)}\sum_{x^u} P(C_1|x^u)x^u}$$

Back to step 1

Why?

$\theta = \{P(C_1), P(C_2), \mu^1, \mu^2, \Sigma\}$

- Maximum likelihood with labelled data Closed-form solution

$$logL(\theta) = \sum_{x^r} logP_\theta(x^r, \hat{y}^r)$$

$$\boxed{P_\theta(x^r, \hat{y}^r) = P_\theta(x^r|\hat{y}^r)P(\hat{y}^r)}$$

- Maximum likelihood with labelled + unlabeled data

$$logL(\theta) = \sum_{x^r} logP_\theta(x^r) + \sum_{x^u} logP_\theta(x^u)$$

$$\boxed{P_\theta(x^u) = P_\theta(x^u|C_1)P(C_1) + P_\theta(x^u|C_2)P(C_2)}$$

Semi-supervised Learning
Low-density Separation

Density: 密度

Low density : the density is low at the boundary

Self-training

- Given: labelled data set = $\{(x^r, \hat{y}^r)\}_{r=1}^{R}$, unlabeled data set = $\{x^u\}_{u=l}^{R+U}$
- Repeat:
  - Train model f^* from labelled data set

    Independent to the model

  - Apply f^* to the unlabeled data set
    - Obtain $\{(x^u, y^u)\}_{u=l}^{R+U}$ Pseudo-label
  - Remove a set of data from unlabeled data set, and add them into ~~un~~labeled data set

How to choose the data set remains open

You can also provide a weight to each data.

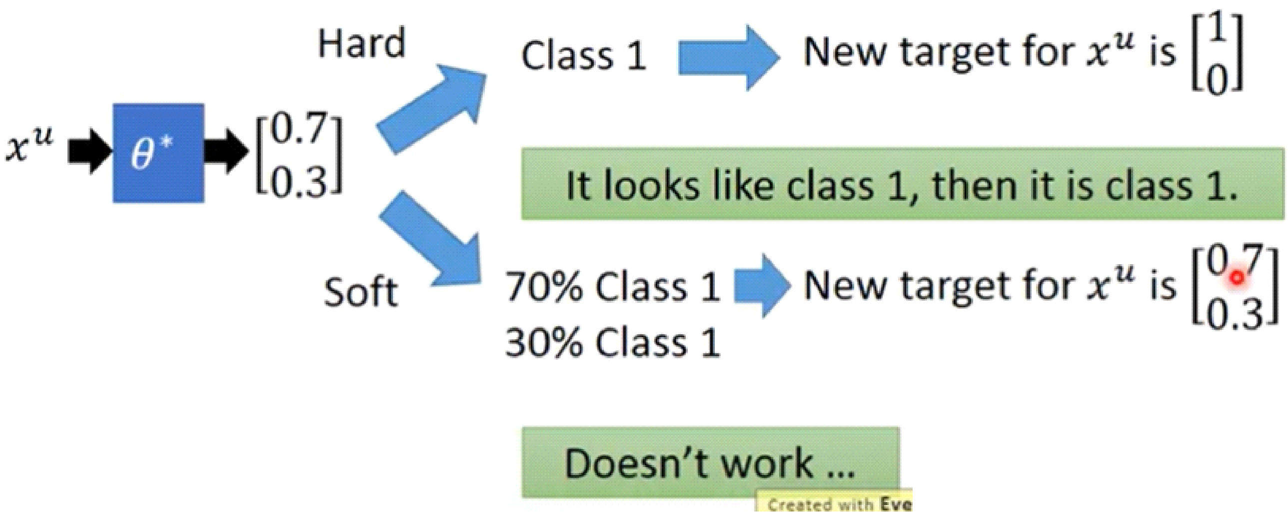
在regression上可能会有用吗?
Output a number.
Train 完 再 train 不会影响 f^star

Self-training

- Similar to semi-supervised learning for generative model
- Hard label v.s. Soft label

Considering using neural network

θ^* (network parameter) from labelled data



Entropy-based Regularization

不符合 low density distribution 的假设

Entropy of y^u :
Evaluate how concentrate the distribution y^u is

$$E(y^u) = -\sum_{m=1}^{5} y_m^u ln(y_m^u)$$

y表示 概率, 前两个都是0
第三个比较大

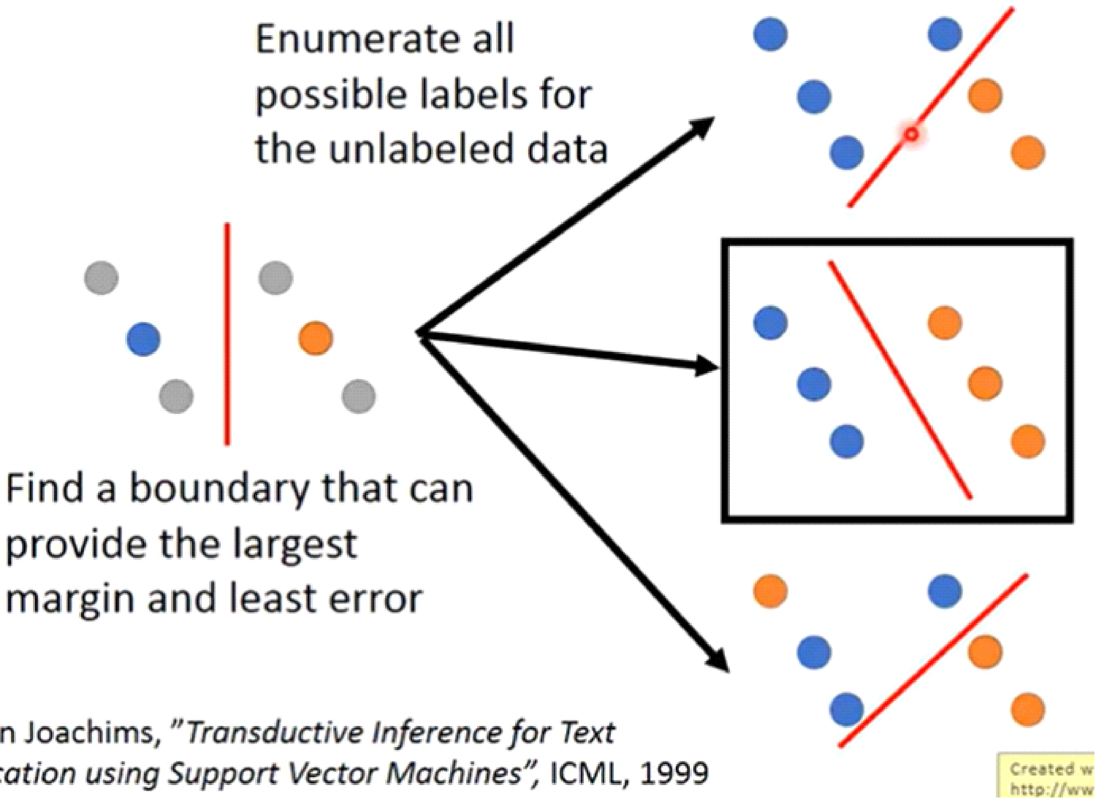
So:
As small as possible

$$L = \sum_{x^r} C(y^r, \hat{y}^r) \quad \text{labelled data}$$

$$+\lambda \sum_{x^u} E(y^u) \quad \text{unlabeled data}$$

Labeled: cross-entropy
Unlabeled: self-cross-entropy(var)
Key: make the distribution as concentrate as possible

Outlook: Semi-supervised SVM



Thorsten Joachims, "*Transductive Inference for Text Classification using Support Vector Machines*", ICML, 1999

Semi-supervised Learning
Smoothness Assumption

近朱者赤，近墨者黑

"You are known by the company you keep"