Asymmetric-noise-based binary classification using EM algorithm.

$$x \to \boxed{\text{classifier}}^{\,w} \to y \xrightarrow{T} \tilde{y}$$

[贝叶斯 MAP 策略]

一. Model

1. confusion matrix T.

   for j-th annotator: $\begin{bmatrix} \alpha^j & 1-\alpha^j \\ 1-\beta^j & \beta^j \end{bmatrix}$.

2. classifier.

$$p(y|x) = \sigma(w^T x).$$

二. Maximum likelihood. $\theta = \{\alpha, \beta, w\}$

$$p(D|\theta) = \prod_{i=1}^{N} p(y_i^1, y_i^2, \dots y_i^R | x_i; \theta).$$

全概率公式

$$= \prod_{i=1}^{N} \{ p(y_i^1, y_i^2, \dots, y_i^R | y=1, x_i; \alpha) \cdot p(y=1|x_i; w) + p(y_i^1, y_i^2, \dots, y_i^R | y=0, x_i; \beta) \cdot p(y=0|x_i; w) \}$$

y_i^j 独立.

$$= \prod_{i=1}^{N} \prod_{j=1}^{R} p(y_i^j | y=1, x_i; \alpha) \cdot p(y=1|x_i; w) + \prod_{j=1}^{R} p(y_i^j | y=0, x_i; \beta) \cdot p(y=0|x_i; w)$$

满足伯努利分布

$$= \prod_{i=1}^{N} \underbrace{\prod_{j=1}^{R} (\alpha^j)^{y_i^j} \cdot (1-\alpha^j)^{1-y_i^j}}_{a_i} \cdot \underbrace{p(y=1|x_i; w)}_{p_i} + \underbrace{\prod_{j=1}^{R} (\beta^j)^{1-y_i^j} \cdot (1-\beta^j)^{y_i^j}}_{b_i} \cdot \underbrace{p(y=0|x_i; w)}_{1-p_i}.$$

$$= \prod_{i=1}^{N} \underbrace{a_i p_i}_{\text{对应}y=1\text{时}} + \underbrace{b_i (1-p_i)}_{\text{对应}y=0\text{时}}.$$

$$\hat{\theta}_{ML} = \arg\max_{\theta} \{ \ln p(D|\theta) \}.$$

三. EM algorithm.

E. 通过对 y 的现有估计值, 计算期望的似然值.

let $\mu_i = p(y_i = 1 | x_i, y_i^1, \dots, y_i^R; \theta)$.

已知 θ, D.

$$\mu_i = \frac{p(y_i^1, \dots y_i^R | y_i = 1, \alpha) \cdot p(y_i = 1 | x, w)}{p(y_i^1, \dots y_i^R | x; \theta)} = \frac{a_i p_i}{a_i p_i + b_i (1-p_i)}$$

M. 最大似求解

$$\theta^{(t+1)} = \arg\max_{\theta} \{ \sum_{i=1}^{n} \mu_i \ln a_i p_i + (1-\mu_i) \ln b_i (1-p_i) \}.$$

$$\alpha^j = \frac{\sum_{i=1}^{n} \mu_i y_i^j}{\sum_{i=1}^{n} \mu_i} \qquad \beta^j = \frac{\sum_{i=1}^{n} (1-\mu_i)(1-y_i^j)}{\sum_{i=1}^{n} (1-\mu_i)}.$$

for w.

$$L(w) = \sum_{i=1}^{n} \left(\mu_i - \sigma(w^T x_i) \right)^2.$$ gradient descent.

总结. (对噪声显式建模 T).

$$\boxed{\quad w \quad} \xrightarrow{y} \boxed{\,T\,} \to \tilde{y}$$

1. generalized EM

$$x\ \boxed{\quad w \quad} \xrightarrow{y} \boxed{\,T\,} \to \tilde{y}$$

估算y后梯度下降　EM 算法.

从 $p(x, \tilde{y})$ 分布中从两边学得 y

2. Annotation confusion. 梯度下降

$$\boxed{\quad w \quad} \xrightarrow{y} \boxed{\,T\,} \to \tilde{y}$$

加入正则化项使得 $T \to$ confusion matrix of each annotator.

3. C-model

$$\boxed{\quad w \quad} \xrightarrow{y} \boxed{\,T\,} \to \tilde{y}$$

将 T 等效为 softmax, 使得 T 从 CM 升格为 $u^T x + b$ 的噪声适应层.