7.6

ノートのタイトル 2016/02/20

$p(t=1|y)=\sigma(y)$

このとき

$p(t=-1|y)=1-p(t=1|y)=1-\sigma(y)=\sigma(-y)$

なので

$p(t|y)=\sigma(yt)$ ‥(7.46)

と書ける。

データの尤度関数は、各データ独立なので

chain rule

$p(t_1,t_2|y_1,y_2)=p(t_1|t_2,y_1,y_2)\,p(t_2|y_1,y_2)$

$=p(t_1|y_1)\,p(t_2|y_2)$

各pから依存しない条件を消した

$p(t_1,t_2\cdots|y_1,y_2\cdots)=p(t_1|y_1)\,p(t_2|y_2)\cdots \underset{(7.46)}{=} \prod_{i=1}^{N}\sigma(y_it_i)$

となる。よって尤度の対数にマイナスつけると

σの定義

$-\ln p(t_1,t_2\cdots|y_1,y_2\cdots)=-\ln\prod_{i=1}^{N}\sigma(y_it_i)=-\sum_{i=1}^{N}\ln\sigma(y_it_i)=-\sum\ln\frac{1}{1+\exp(-y_it_i)}=\sum\ln(1+\exp(-y_it_i))$

よって誤差関数は

$E=\sum\ln(1+\exp(-y_it_i))$

正則化項をつけて

$E=\sum\ln(1+\exp(-y_it_i))+\lambda\|w\|^2$

ここで

$E_{LR}(yt)=\ln(1+\exp(-yt))$ ‥(7.48)

とおくと

$E=\sum_{i=1}^{N}E_{LR}(y_it_i)+\lambda\|w\|^2$ ‥(7.47)

を得る