Building a Discriminative Generative Model

Ruei-sung Lin David Ross Jongwoo Lim Ming-Hsuan Yang

Feb 14, 2004

Denote $P(y|C) \sim \mathcal{N}(0, C)$ the probability of generative y from the model. Now given another set of examples $Y = (y_1, \ldots, y_M)$ that are known not coming from the model, we want to find a linear projection W that projects observation y into a r-dimensional subspace such that the likelihood for the model to gererating examples WY in the subspace is *minimized.*

Since $P(y|C)$ is a Gaussian, the linear project of y is also a Gaussian $Wy \sim \mathcal{N}(0, WCW^T)$

The log likelihood $\mathcal{L}(WY)$ is

$$\begin{aligned} L(WY) &= \sum_{i=1}^{M} -\frac{1}{2}\left(r\ln(2\pi) + \ln|WCW^T| + y_i^T(WCW^T)^{-1}y_i\right) \\ &= -\frac{M}{2}\left(r\ln(2\pi) + \ln|WCW^T| + \frac{1}{M}\sum_{i=1}^{M} y_i^T W^T (WCW^T)^{-1} W y_i\right) \\ &= -\frac{M}{2}\left(r\ln(2\pi) + \ln|WCW^T| + tr\left((WCW^T)^{-1} W S_W W^T\right)\right) \end{aligned}$$

with $S_W = \frac{1}{M}\sum_{i=1}^{N} y_i^- y_i^{-T}$. We first consider the case when $r = 1$, i.e. $W = w^T$

$$tr((w^T C w)^{-1} w^T S_W w) = \frac{w^T S_W w}{w^T C w}$$

and $\mathcal{L}(wY^-)$ becomes

$$\mathcal{L}(w^T Y^-) = -\frac{M}{2}\left\{r\ln(2\pi) + \ln|w^T C w| + \frac{w^T S_W w}{w^T C w}\right\}$$

minimize $L(w^T Y^-)$ is equivalent to maximize $\mathcal{L}'(w)$

$$\mathcal{L}'(w) = \ln|w^T C w| + \frac{w^T S_W w}{w^T C w}$$

Without constraints $\mathcal{L}'(w)$, increasing the length of the vector, i.e. $|w|_2$, will make the value of $\mathcal{L}'(w)$ grow toward ∞. To avoid this, we impose a constraint, $w^T Cw = 1$. This becomes a constraint optimization problem:

$$w^* = \arg\max_w \phi(w) = w^T S_W w$$

under the constraint,

$$w^T Cw = 1$$

w^* can then be obtained by solving the generalized eigenvalue problem: $eig(w^T S_W w, w^T Cw)$. The extension to cases when $r > 1$ is straight-forward.