

Lec 8. HW1

$$J(\mu, \Sigma) = (\max(0, 1 - \mu^T \phi(x) y))^2 + \phi(x)^T \Sigma \phi(x) + r \left\{ \log \frac{\det(\tilde{\Sigma})}{\det(\Sigma)} + \text{tr}(\tilde{\Sigma}^{-1} \Sigma) + (\mu - \tilde{\mu})^T \tilde{\Sigma}^{-1} (\mu - \tilde{\mu}) - d \right\}$$

$$\hat{\mu} = \arg \min_{\mu} J(\mu, \Sigma) \quad \text{is the goal.}$$

$$\frac{\partial}{\partial \mu} J(\mu, \Sigma) = \frac{\partial}{\partial \mu} \underbrace{[(\max(0, 1 - \mu^T \phi(x) y))^2]}_{(1)} + 2r \tilde{\Sigma}^{-1} (\mu - \tilde{\mu})$$

$$(1) \text{ is the goal.}$$

$$1 - \mu^T \phi(x) y > 0 \quad \text{a.s.}$$

$$\begin{aligned} (1) \rightarrow \frac{\partial}{\partial \mu} (1 - \mu^T \phi(x) y)^2 &= \frac{\partial}{\partial \mu} \|1 - \mu^T \phi(x) y\|^2 \quad \left. \begin{array}{l} \because (1 - \mu^T \phi(x) y) \in \mathbb{R} \\ = -2 \phi(x) y (1 - \mu^T \phi(x) y) \\ = -2 \phi(x) y (1 - y \phi(x)^T \mu) \end{array} \right\} \text{ s.t.} \end{aligned}$$

$$\frac{\partial}{\partial \mu} J = 0$$

$$\rightarrow -\phi(x) y + \phi(x) \overset{1}{\underset{1}{y}}^2 \phi(x)^T \hat{\mu} + r \tilde{\Sigma}^{-1} (\hat{\mu} - \tilde{\mu}) = 0$$

$$\rightarrow (\phi(x) \phi(x)^T + r \tilde{\Sigma}^{-1}) \hat{\mu} = \phi(x) y + r \tilde{\Sigma}^{-1} \tilde{\mu}$$

$$B = \phi(x)$$

$$D = \frac{y^2}{2} I$$

$$C = \phi(x)^T$$

$$\rightarrow \hat{\mu} = (r \tilde{\Sigma}^{-1} + \phi(x) \phi(x)^T)^{-1} (r \tilde{\Sigma}^{-1} \tilde{\mu} + \phi(x) y)$$

$$= (\tilde{\Sigma}^{-1} + \frac{1}{r} \phi(x) \phi(x)^T)^{-1} (\tilde{\Sigma}^{-1} \tilde{\mu} + \frac{y}{r} \phi(x)) \quad \leftarrow \text{Woodbury matrix identity}$$

$$= \left\{ \tilde{\Sigma} - \tilde{\Sigma} \phi(x) \left[\frac{1}{r} + \phi(x)^T \tilde{\Sigma} \phi(x) \right]^{-1} \phi(x)^T \tilde{\Sigma} \right\} (\tilde{\Sigma}^{-1} \tilde{\mu} + \frac{y}{r} \phi(x))$$

$$= \left(\tilde{\Sigma} - \frac{\tilde{\Sigma} \phi(x) \phi(x)^T \tilde{\Sigma}}{\phi(x)^T \tilde{\Sigma} \phi(x) + r} \right) (\tilde{\Sigma}^{-1} \tilde{\mu} + \frac{y}{r} \phi(x))$$

$$= \tilde{\mu} + \frac{y}{r} \tilde{\Sigma} \phi(x) - \frac{\tilde{\Sigma} \phi(x) \phi(x)^T \tilde{\mu} + \tilde{\Sigma} \phi(x) \phi(x)^T \tilde{\Sigma} \phi(x) \cdot \frac{y}{r}}{\phi(x)^T \tilde{\Sigma} \phi(x) + r}$$

$$= \tilde{\mu} - \frac{\tilde{\Sigma} \phi(x) \phi(x)^T \tilde{\mu}}{\phi(x)^T \tilde{\Sigma} \phi(x) + r} - \frac{y \tilde{\Sigma} \phi(x)}{\phi(x)^T \tilde{\Sigma} \phi(x) + r} \quad \text{7th line}$$

$$= \tilde{\mu} + \frac{y \tilde{\Sigma} \phi(x) - \tilde{\mu} \phi(x)^T \tilde{\Sigma} \phi(x)}{\phi(x)^T \tilde{\Sigma} \phi(x) + r}$$

$$= \tilde{\mu} + \frac{y (1 - \mu^T \phi(x) y)}{\phi(x)^T \tilde{\Sigma} \phi(x) + r} \tilde{\Sigma} \phi(x)$$

$$\left(\frac{y}{r} < 1 \right)$$

... (A)