1. (Math) Suppose for multiclass logistic regression

$$p(x|y=i) = N(x|\mu_i, \Sigma),$$

Compute

$$a_i = \ln[p(x|y=i)p(y=i)].$$

Is it still linear? What if $p(x|y=i)$ has different covariance? That is,

$$p(x|y=i) = N(x|\mu_i, \Sigma_i),$$

$\longrightarrow$ Multivariate Gaussian Distribution

$$N(x|\mu, \Sigma) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right\}$$

$$= \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left\{\frac{(x-\mu)^T(x-\mu)}{-2\Sigma}\right\}$$

1) for $p(x|y=i) = N(x|\mu_i, \Sigma)$

$$a_i = \ln[p(x|y=i)p(y=i)]$$

$$= \ln[p(x|y=i)] + \ln[p(y=i)]$$

$$= \ln[N(x|\mu_i, \Sigma)] + \ln[p(y=i)]$$

$$= \ln\left[\frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu_i)^T\Sigma^{-1}(x-\mu_i)\right)\right] + \ln[p(y=i)]$$

$$= -\frac{1}{2}\left[\ln(|\Sigma|) + (x-\mu_i)^T\Sigma^{-1}(x-\mu_i) + d\ln(2\pi)\right] + \ln(p(y=i))$$

$$= -\frac{1}{2}(x-\mu_i)^T\Sigma^{-1}(x-\mu_i) - \frac{1}{2}\left[\ln(|\Sigma|) + d\ln(2\pi)\right] + \ln(p(y=i))$$

$$= -\frac{1}{2\Sigma}\left(x^Tx - 2\mu_i x + \mu_i^T\mu_i\right) - \frac{1}{2}\left[\ln(|\Sigma|) + d\ln(2\pi)\right] + \ln(p(y=i))$$

$$= \underbrace{-\frac{1}{2\Sigma}x^Tx}_{\text{Cancel out}} + \underbrace{\frac{1}{\Sigma}\mu_i \cdot x}_{w^Tx} \underbrace{- \frac{1}{2\Sigma}\mu_i^T\mu_i - \frac{1}{2}\left[\ln(|\Sigma|) + d\ln(2\pi)\right] + \ln(p(y=i))}_{\text{Bias}}$$

$\Rightarrow$ Linear where $\begin{cases} w = \frac{1}{\Sigma}\mu_i \\ b = -\frac{1}{2}\left[\Sigma^{-1}\mu_i^T\mu_i + \ln|\Sigma| + d\ln 2\pi\right] + \ln p(y=i) \end{cases}$

2) for $p(x|y=i) = \mathcal{N}(x|\mu_i, \Sigma_i)$

$$a_i = \ln[p(x|y=i)p(y=i)]$$

$$= \ln[\mathcal{N}(x|\mu_i, \Sigma_i)] + \ln[p(y=i)]$$

$$= \ln\left[\frac{1}{(2\pi)^{d/2}|\Sigma_i|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu_i)^T \Sigma_i^{-1}(x-\mu_i)\right)\right] + \ln[p(y=i)]$$

$$= -\frac{1}{2}\left[\ln(|\Sigma_i|) + (x-\mu_i)^T \Sigma_i^{-1}(x-\mu_i) + d\ln(2\pi)\right] + \ln(p(y=i))$$

$$= -\frac{1}{2}(x-\mu_i)^T \Sigma_i^{-1}(x-\mu_i) - \frac{1}{2}\left[\ln(|\Sigma_i|) + d\ln(2\pi)\right] + \ln(p(y=i))$$

$$= \underbrace{-\frac{1}{2\Sigma_i}x^T x}_{\text{Canceled out}} + \underbrace{\frac{1}{\Sigma_i}\mu_i x}_{w^T x} \underbrace{- \frac{1}{2\Sigma_i}\mu_i^T \mu_i - \frac{1}{2}\ln|\Sigma_i| - \frac{1}{2}d\ln 2\pi + \ln p(y=i)}_{\text{bias}}$$

also linear where $\begin{cases} w = \Sigma_i^{-1}\mu_i \\ b = -\frac{1}{2}\left[\Sigma_i^{-1}\mu_i^T \mu_i + \ln|\Sigma_i| + d\ln 2\pi\right] + \ln p(y=i) \end{cases}$

2. (Math) Recall that the SVM objective can be simplified as minimizing

$$\tilde{L}(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j$$

subject to $\alpha_i \geq 0 (\forall i)$, and $\sum_{i=1}^{n} \alpha_i y_i = 0$. Suppose the optimal solution is $\alpha^* = (\alpha_1^*, \alpha_2^*, \cdots, \alpha_n^*)$. Show that the margin γ satisfies

$$\frac{1}{\gamma^2} = \sum_{i=1}^{n} \alpha_i^*$$

$$RHS = \sum_{i=1}^{n} \alpha_i^*$$

$$= \tilde{L}(\alpha^*) + \frac{1}{2}\sum_i \sum_j \alpha_i \alpha_j y_i y_j x_i^T x_j$$

$$LHS = \frac{1}{\gamma^2} = \min \|w\|^2 \quad \because \gamma = \frac{1}{\|w\|}$$

$$= \min \frac{1}{2}\|w\|^2 + \min \frac{1}{2}\|w\|^2$$

$$= \tilde{L}(\alpha^*) + \frac{1}{2}\sum_i \sum_j \alpha_i \alpha_j y_i y_j x_i^T x_j$$

Notes:

① With optimal Lagrange multiplier α^*, Lagrange have 0 penalty term

$\hookrightarrow L(\alpha^*) = \frac{1}{2}\|w\|^2$

② $\frac{\partial L}{\partial w} = 0 \Rightarrow w = \sum_i \alpha_i y_i x_i$

plug into $\frac{1}{2}\|w\|^2$

$\hookrightarrow \frac{1}{2}\sum_{ij} \alpha_i \alpha_j y_i y_j x_i^T x_j$

$$\boxed{LHS = RHS}$$

Given $\frac{\partial L}{\partial w} = 0 \Rightarrow w = \sum_i \alpha_i y_i x_i$ ①

$\frac{\partial L}{\partial b} = 0 \Rightarrow 0 = \sum_i \alpha_i y_i$ ②

$$L(w, b, \alpha) = \frac{1}{2}\|w\|^2 - \sum_i \alpha_i [y_i(w^T x_i + b) - 1]$$

$$= \frac{1}{2}\|w\|^2 - \sum_i \alpha_i y_i w^T x_i + \alpha_i y_i b - \alpha_i y_i$$

$$= \frac{1}{2}\sum_i \alpha_i y_i x_i \sum_j \alpha_j y_j x_j$$

$$- \sum_i \alpha_i y_i x_i \sum_j \alpha_j y_j x_j + \overset{0}{\cancel{\sum_i \alpha_i y_i}} b + \sum_i \alpha_i$$

$$= \frac{1}{2}\sum_{ij} \alpha_i \alpha_j y_i y_j x_i^T x_j - \sum_{ij} \alpha_i \alpha_j y_i y_j x_i^T x_j$$

$$+ \sum_i \alpha_i$$

$$= \sum_i \alpha_i - \frac{1}{2}\sum_{ij} \alpha_i \alpha_j y_i y_j x_i^T x_j$$