Gradient Descent Optimization

Function to maximize:

$$\hat{h} = \arg\max_{h \in \mathcal{H}} P(S|h)P(h) \tag{1}$$

Then: Naive Bayes:

$$\hat{h} = \arg\max_{h \in \mathcal{H}} P(h) \prod_{d=1}^{D} P(s_d|h) \tag{2}$$

Then: since h is composed by K-zones:

$$\hat{h} = \arg\max_{h \in \mathcal{H}} \prod_{k=1}^{K} \prod_{d=1}^{|S_t|} P(s_d|h_k)P(h_k) \tag{3}$$

where S_k is the sub-set of sites that belongs into the layout zone k.
From here we work under the constrain of a single main zone defined by its corners $(\boldsymbol{u}, \boldsymbol{b})$
Then: $P(h_k)$ is divided in its components:

$$\hat{h} = \arg\max_{h \in \mathcal{H}} \prod_{k=1}^{K} \prod_{d=1}^{|S_k|} P(s_d|(\boldsymbol{u}_k, \boldsymbol{b}_k))P(\boldsymbol{u}_k, \boldsymbol{b}_k) \tag{4}$$

Then: applying log:

$$\begin{aligned} \log \hat{h} &= \arg\max_{h \in \mathcal{H}} \log \prod_{k=1}^{K} \prod_{d=1}^{|S_k|} P(s_d|(\boldsymbol{u}_k, \boldsymbol{b}_k))P(\boldsymbol{u}_k, \boldsymbol{b}_k) \\ &= \arg\max_{h \in \mathcal{H}} \sum_{k=1}^{K} \sum_{d=1}^{|S_k|} \log P(s_d|(\boldsymbol{u}_k, \boldsymbol{b}_k)) + \log P(\boldsymbol{u}_k, \boldsymbol{b}_k) \end{aligned} \tag{5}$$

Then $P(\boldsymbol{u}_k, \boldsymbol{b}_k)$ is modeled by a Gaussian Mixture, and under independence assumption of $\boldsymbol{u}$ and $\boldsymbol{b}$, but restricted to $\boldsymbol{u} < \boldsymbol{b}$ (element wise):

restriction only for programming, there is no mathematical reason behind

$$P(\boldsymbol{u}_k, \boldsymbol{b}_k) = P(\boldsymbol{u}_k)P(\boldsymbol{b}_k) = \sum_{g=1}^{G_{\boldsymbol{u}_k}} \phi_g \mathcal{N}(\boldsymbol{u}_k, \boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g) \sum_{g=1}^{G_{\boldsymbol{b}_k}} \phi_g \mathcal{N}(\boldsymbol{b}_k, \boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g) \tag{6}$$

Then: using *log*

$$\log P(\boldsymbol{u}_k)P(\boldsymbol{b}_k) = \log \sum_{g=1}^{G_{\boldsymbol{u}_k}} \phi_g \mathcal{N}(\boldsymbol{u}_k, \boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g) + \log \sum_{g=1}^{G_{\boldsymbol{b}_k}} \phi_g \mathcal{N}(\boldsymbol{b}_k, \boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g) \quad (7)$$

then: derivative respect to $\boldsymbol{u}$ and $\boldsymbol{b}$:

$$\frac{\delta \log P(\boldsymbol{u_k})P(\boldsymbol{b_k})}{\delta \boldsymbol{u}_k, \boldsymbol{b}_k} = \frac{\delta}{\delta \boldsymbol{u_k}} \log \sum_{g=1}^{G_{\boldsymbol{u}_k}} \phi_g \mathcal{N}(\boldsymbol{u}_k, \boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g) + \frac{\delta}{\delta \boldsymbol{b}_k} \log \sum_{g=1}^{G_{\boldsymbol{b}_k}} \phi_g \mathcal{N}(\boldsymbol{b}_k, \boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g) \quad (8)$$

Then, only for $\boldsymbol{u}_k$ dependent term:

$$= \frac{\delta}{\delta \boldsymbol{u_k}} \log \sum_{g=1}^{G_{\boldsymbol{u}_k}} \phi_g \mathcal{N}(\boldsymbol{u}_k, \boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)$$

Using: $\frac{\delta \log f(x)}{\delta x} = \frac{1}{f(x)} \frac{\delta f(x)}{\delta x}$

$$= \frac{1}{\sum_{g=1}^{G_{\boldsymbol{u}_k}} \phi_g \mathcal{N}(\boldsymbol{u}_k, \boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)} \frac{\delta}{\delta \boldsymbol{u}_k} \sum_{g=1}^{G_{\boldsymbol{u}_k}} \phi_g \mathcal{N}(\boldsymbol{u}_k, \boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)$$

Replace GMM definition and taking out independent terms

$$= \frac{1}{\sum_{g=1}^{G_{\boldsymbol{u}_k}} \phi_g \mathcal{N}(\boldsymbol{u}_k, \boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)} \Sigma_{g=1}^{G_{\boldsymbol{u}_k}} \phi_g (2\pi)^{-G_{\boldsymbol{u}_k}/2} |\boldsymbol{\Sigma}_g|^{-1/2} \frac{\delta}{\delta \boldsymbol{u}_k} \exp^{-1/2(\boldsymbol{u}_k - \boldsymbol{\mu}_g)^T \boldsymbol{\Sigma}_g^{-1} (\boldsymbol{u}_k - \boldsymbol{\mu}_g)}$$

Derivative exponential term

$$= \frac{\Sigma_{g=1}^{G_{\boldsymbol{u}_k}} \phi_g \overbrace{(2\pi)^{-G_{\boldsymbol{u}_k}/2} |\boldsymbol{\Sigma}_g|^{-1/2} \exp^{-1/2(\boldsymbol{u}_k - \boldsymbol{\mu}_g)^T \boldsymbol{\Sigma}_g^{-1} (\boldsymbol{u}_k - \boldsymbol{\mu}_g)}}^{\mathcal{N}(\boldsymbol{u}_k, \boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)} \frac{\delta}{\delta \boldsymbol{u}_k} (-1/2(\boldsymbol{u}_k - \boldsymbol{\mu}_g)^T \boldsymbol{\Sigma}_g^{-1} (\boldsymbol{u}_k - \boldsymbol{\mu}_g))}{\sum_{g=1}^{G_{\boldsymbol{u}_k}} \phi_g \mathcal{N}(\boldsymbol{u}_k, \boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)}$$

Using Lemma 6.2.3 for symmetric $\boldsymbol{\Sigma_g^{-1}}$ [1]

$$= \frac{-\Sigma_{g=1}^{G_{\boldsymbol{u}_k}} \phi_g \mathcal{N}(\boldsymbol{u}_k, \boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)(\boldsymbol{u}_k - \boldsymbol{\mu}_g)^T \boldsymbol{\Sigma_g^{-1}}}{\sum_{g=1}^{G_{\boldsymbol{u}_k}} \phi_g \mathcal{N}(\boldsymbol{u}_k, \boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)} \quad (9)$$

Now from Eq. 9 in Eq. 8:

$$
\begin{aligned}
\frac{\delta \log P(\boldsymbol{u_k})P(\boldsymbol{b_k})}{\delta \boldsymbol{u}_k, \boldsymbol{b}_k} &= \frac{-\Sigma_{g=1}^{G_{\boldsymbol{u}_k}} \phi_g \mathcal{N}(\boldsymbol{u}_k, \boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)(\boldsymbol{u}_k - \boldsymbol{\mu}_g)^T \boldsymbol{\Sigma_g^{-1}}}{\sum_{g=1}^{G_{\boldsymbol{u}_k}} \phi_g \mathcal{N}(\boldsymbol{u}_k, \boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)} \\
&+ \frac{-\Sigma_{g=1}^{G_{\boldsymbol{b}_k}} \phi_g \mathcal{N}(\boldsymbol{b}_k, \boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)(\boldsymbol{b}_k - \boldsymbol{\mu}_g)^T \boldsymbol{\Sigma_g^{-1}}}{\sum_{g=1}^{G_{\boldsymbol{b}_k}} \phi_g \mathcal{N}(\boldsymbol{b}_k, \boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)} \qquad (10)
\end{aligned}
$$

References

[1] Luo, Y. Local Gradient Descent Methods for GMM Simplification. Tech. rep., 2015.