

Gradient Descent Optimization

Function to maximize:

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

Then: Naive Bayes:

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

Then: since h is composed by K-zones:

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

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 (\mathbf{u}, \mathbf{b})

Then: $P(h_k)$ is divided in its components:

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

Then: applying \log :

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

Now we want to search for the best $(\mathbf{u}_k^{i+1}, \mathbf{b}_k^{i+1})$ after some $(\mathbf{u}_k^i, \mathbf{b}_k^i)$, in order to do that we can use gradient descent optimization over Eq. 5.

First for \mathbf{u}_k :

$$\begin{aligned}
\mathbf{u}_k^{i+1} &= \mathbf{u}_k^i - \alpha \frac{\delta h}{\delta \mathbf{u}_k^i} \\
&= \mathbf{u}_k^i - \alpha \frac{\delta}{\delta \mathbf{u}_k^i} \sum_{k=1}^K \sum_{d=1}^{|S_k|} \log P(s_d | (\mathbf{u}_k, \mathbf{b}_k)) + \log P(\mathbf{u}_k, \mathbf{b}_k) \\
&= \mathbf{u}_k^i - \alpha \sum_{k=1}^K \sum_{d=1}^{|S_k|} \left(\overbrace{\frac{\delta}{\delta \mathbf{u}_k^i} \log P(s_d | (\mathbf{u}_k, \mathbf{b}_k))}^{\gamma} + \underbrace{\frac{\delta}{\delta \mathbf{u}_k^i} \log P(\mathbf{u}_k, \mathbf{b}_k)}_{\beta} \right) \quad (6)
\end{aligned}$$

Now, we can take γ and β separately:

For β :

$$\begin{aligned}
\beta &= \frac{\delta}{\delta \mathbf{u}_k^i} \log P(\mathbf{u}_k^i) P(\mathbf{b}_k^i) \\
&= \frac{\delta}{\delta \mathbf{u}_k^i} \log \sum_{g=1}^{G_{\mathbf{u}_k^i}} \phi_g \mathcal{N}(\mathbf{u}_k^i, \boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g) + \frac{\delta}{\delta \mathbf{u}_k^i} \log \sum_{g=1}^{G_{\mathbf{b}_k^i}} \phi_g \mathcal{N}(\mathbf{b}_k^i, \boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g) \xrightarrow{0}
\end{aligned}$$

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

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

Using: (where GMM is defined)

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

Using: $\frac{\delta \exp^{f(x)}}{\delta x} = \exp^{f(x)} \frac{\delta f(x)}{\delta x}$

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

Using: Lemma 6.2.3 for symmetric Σ_g^{-1} [1]

$$\begin{aligned}
&= D \sum_{g=1}^{G_{\mathbf{u}_k^i}} \phi_g \mathcal{N}(\mathbf{u}_k^i, \boldsymbol{\mu}_g, \Sigma_g) (\mathbf{u}_k^i - \boldsymbol{\mu}_g)^T \Sigma_g^{-1} \\
&= \frac{\sum_{g=1}^{G_{\mathbf{u}_k^i}} \phi_g \mathcal{N}(\mathbf{u}_k^i, \boldsymbol{\mu}_g, \Sigma_g) (\mathbf{u}_k^i - \boldsymbol{\mu}_g)^T \Sigma_g^{-1}}{\sum_{g=1}^{G_{\mathbf{u}_k^i}} \phi_g \mathcal{N}(\mathbf{u}_k^i, \boldsymbol{\mu}_g, \Sigma_g)} \tag{7}
\end{aligned}$$

Now, for γ ; due nature γ an analytic solution is very hard to obtain, so instead a geometric approach is followed.

Using one dimensional Five-Points Stencil ¹ method on each axis we can obtain the first derivative on the point $(\mathbf{u}_k^i, \mathbf{b}_k^i)$:

$$\begin{aligned}
\gamma_r = & \frac{-\log P(s_d | ((r_{\mathbf{u}_k^i} + 2\Delta r, c_{\mathbf{u}_k^i}), \mathbf{b}_k^i))}{12\Delta r} \\
& + \frac{8 \log P(s_d | ((r_{\mathbf{u}_k^i} + \Delta r, c_{\mathbf{u}_k^i}), \mathbf{b}_k^i))}{12\Delta r} \\
& - \frac{8 \log P(s_d | ((r_{\mathbf{u}_k^i} - \Delta r, c_{\mathbf{u}_k^i}), \mathbf{b}_k^i))}{12\Delta r} \\
& + \frac{\log P(s_d | ((r_{\mathbf{u}_k^i} - 2\Delta r, c_{\mathbf{u}_k^i}), \mathbf{b}_k^i))}{12\Delta r} \tag{8}
\end{aligned}$$

$$\begin{aligned}
\gamma_c = & \frac{-\log P(s_d | ((r_{\mathbf{u}_k^i}, c_{\mathbf{u}_k^i} + 2\Delta c), \mathbf{b}_k^i))}{12\Delta r} \\
& + \frac{8 \log P(s_d | ((r_{\mathbf{u}_k^i}, c_{\mathbf{u}_k^i} + \Delta c), \mathbf{b}_k^i))}{12\Delta r} \\
& - \frac{8 \log P(s_d | ((r_{\mathbf{u}_k^i}, c_{\mathbf{u}_k^i} - \Delta c), \mathbf{b}_k^i))}{12\Delta r} \\
& + \frac{\log P(s_d | ((r_{\mathbf{u}_k^i}, c_{\mathbf{u}_k^i} - 2\Delta c), \mathbf{b}_k^i))}{12\Delta r} \tag{9}
\end{aligned}$$

¹ $f'(x) \approx \frac{-f(x+2h)+8f(x+h)-8f(x-h)+f(x-2h)}{12h}$; h = space between points in the grid

References

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