Yke Rusticus
11306386
yke.rusticus@student.uva.nl

1 Mixture of Experts

1.1

$$p(\mathbf{y} \mid \mathbf{X}, \mathbf{\Theta}, \mathbf{\Phi}) = \prod_{n=1}^{N} \sum_{k=1}^{K} p(z_n = k \mid \mathbf{x}_n, \mathbf{\Phi}) p(y_n \mid \mathbf{x}_n, \boldsymbol{\theta}_k, z_n = k) \tag{1}$$

$$= \prod_{n=1}^{N} \sum_{k=1}^{K} \pi_{nk} \cdot \text{Exponential}(y_n \mid \lambda = \exp(\boldsymbol{\theta}_k^\top \mathbf{x}_n)) \tag{2}$$

$$\ln p(\mathbf{y} \mid \mathbf{X}, \mathbf{\Theta}, \mathbf{\Phi}) = \sum_{n=1}^{N} \ln \sum_{k=1}^{K} \pi_{nk} \cdot \text{Exponential}(y_n \mid \lambda = \exp(\boldsymbol{\theta}_k^\top \mathbf{x}_n)) \tag{3}$$

1.2

Responsibility of expert i for datapoint n:

$$r_{ni} = p(z_n = i \mid \mathbf{x}_n, y_n) \tag{4}$$

posterior = likelihood $\times$ prior / evidence, so if we interpret the given routing mechanism as a prior over z_n, then

$$r_{ni} = \frac{p(y_n \mid \mathbf{x}_n, \boldsymbol{\theta}_i, z_n = i) p(z_n = i \mid \mathbf{x}_n, \mathbf{\Phi})}{\sum_{k=1}^{K} p(y_n \mid \mathbf{x}_n, \boldsymbol{\theta}_k, z_n = k) p(z_n = k \mid \mathbf{x}_n, \mathbf{\Phi})} \tag{5}$$

$$= \frac{\pi_{ni} \cdot \text{Exponential}(y_n \mid \lambda = \exp(\boldsymbol{\theta}_i^\top \mathbf{x}_n))}{\sum_{k=1}^{K} \pi_{nk} \cdot \text{Exponential}(y_n \mid \lambda = \exp(\boldsymbol{\theta}_k^\top \mathbf{x}_n))} \tag{6}$$

1.3

$$\frac{\partial}{\partial \boldsymbol{\theta}_i} \ln p(\mathbf{y} \mid \mathbf{X}, \mathbf{\Theta}, \mathbf{\Phi}) = \frac{\partial}{\partial \boldsymbol{\theta}_i} \sum_{n=1}^{N} \ln \sum_{k=1}^{K} p(z_n = k \mid \mathbf{x}_n, \mathbf{\Phi}) p(y_n \mid \mathbf{x}_n, \boldsymbol{\theta}_k, z_n = k) \tag{7}$$

$$= \sum_{n=1}^{N} \frac{\partial}{\partial \boldsymbol{\theta}_i} \ln \sum_{k=1}^{K} p(z_n = k \mid \mathbf{x}_n, \mathbf{\Phi}) p(y_n \mid \mathbf{x}_n, \boldsymbol{\theta}_k, z_n = k) \tag{8}$$

$$= \sum_{n=1}^{N} \frac{\frac{\partial}{\partial \boldsymbol{\theta}_i} \sum_{k=1}^{K} p(z_n = k \mid \mathbf{x}_n, \mathbf{\Phi}) p(y_n \mid \mathbf{x}_n, \boldsymbol{\theta}_k, z_n = k)}{\sum_{k=1}^{K} p(z_n = k \mid \mathbf{x}_n, \mathbf{\Phi}) p(y_n \mid \mathbf{x}_n, \boldsymbol{\theta}_k, z_n = k)} \tag{9}$$

$$= \sum_{n=1}^{N} \frac{\frac{\partial}{\partial \boldsymbol{\theta}_i} p(z_n = i \mid \mathbf{x}_n, \mathbf{\Phi}) p(y_n \mid \mathbf{x}_n, \boldsymbol{\theta}_i, z_n = i)}{\sum_{k=1}^{K} p(z_n = k \mid \mathbf{x}_n, \mathbf{\Phi}) p(y_n \mid \mathbf{x}_n, \boldsymbol{\theta}_k, z_n = k)} \tag{10}$$

$$= \sum_{n=1}^{N} \frac{p(z_n = i \mid \mathbf{x}_n, \mathbf{\Phi}) \frac{\partial}{\partial \boldsymbol{\theta}_i} p(y_n \mid \mathbf{x}_n, \boldsymbol{\theta}_i, z_n = i)}{\sum_{k=1}^{K} p(z_n = k \mid \mathbf{x}_n, \mathbf{\Phi}) p(y_n \mid \mathbf{x}_n, \boldsymbol{\theta}_k, z_n = k)} \tag{11}$$

$$= \sum_{n=1}^{N} \frac{p(z_n = i \mid \mathbf{x}_n, \mathbf{\Phi}) p(y_n \mid \mathbf{x}_n, \boldsymbol{\theta}_i, z_n = i)}{\sum_{k=1}^{K} p(z_n = k \mid \mathbf{x}_n, \mathbf{\Phi}) p(y_n \mid \mathbf{x}_n, \boldsymbol{\theta}_k, z_n = k)} \frac{\frac{\partial}{\partial \boldsymbol{\theta}_i} p(y_n \mid \mathbf{x}_n, \boldsymbol{\theta}_i, z_n = i)}{p(y_n \mid \mathbf{x}_n, \boldsymbol{\theta}_i, z_n = i)} \tag{12}$$

$$= \sum_{n=1}^{N} r_{ni} \frac{\partial}{\partial \boldsymbol{\theta}_i} \ln p(y_n \mid \mathbf{x}_n, \boldsymbol{\theta}_i, z_n = i) \tag{13}$$

$$\frac{\partial}{\partial \boldsymbol{\phi}_i} \ln p(\mathbf{y} \mid \mathbf{X}, \boldsymbol{\Theta}, \boldsymbol{\Phi}) = \frac{\partial}{\partial \boldsymbol{\phi}_i} \sum_{n=1}^{N} \ln \sum_{k=1}^{K} p(z_n = k \mid \mathbf{x}_n, \boldsymbol{\Phi}) p(y_n \mid \mathbf{x}_n, \boldsymbol{\theta}_k, z_n = k) \tag{14}$$

$$= \sum_{n=1}^{N} \frac{\partial}{\partial \boldsymbol{\phi}_i} \ln \sum_{k=1}^{K} p(z_n = k \mid \mathbf{x}_n, \boldsymbol{\Phi}) p(y_n \mid \mathbf{x}_n, \boldsymbol{\theta}_k, z_n = k) \tag{15}$$

$$= \sum_{n=1}^{N} \frac{\frac{\partial}{\partial \boldsymbol{\phi}_i} \sum_{k=1}^{K} p(z_n = k \mid \mathbf{x}_n, \boldsymbol{\Phi}) p(y_n \mid \mathbf{x}_n, \boldsymbol{\theta}_k, z_n = k)}{\sum_{k=1}^{K} p(z_n = k \mid \mathbf{x}_n, \boldsymbol{\Phi}) p(y_n \mid \mathbf{x}_n, \boldsymbol{\theta}_k, z_n = k)} \tag{16}$$

$$= \sum_{n=1}^{N} \frac{\sum_{k=1}^{K} p(y_n \mid \mathbf{x}_n, \boldsymbol{\theta}_k, z_n = k) \frac{\partial}{\partial \boldsymbol{\phi}_i} p(z_n = k \mid \mathbf{x}_n, \boldsymbol{\Phi})}{\sum_{k=1}^{K} p(z_n = k \mid \mathbf{x}_n, \boldsymbol{\Phi}) p(y_n \mid \mathbf{x}_n, \boldsymbol{\theta}_k, z_n = k)} \tag{17}$$

$$= \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} \frac{\frac{\partial}{\partial \boldsymbol{\phi}_i} p(z_n = k \mid \mathbf{x}_n, \boldsymbol{\Phi})}{p(z_n = k \mid \mathbf{x}_n, \boldsymbol{\Phi})} \tag{18}$$

$$= \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} \frac{\partial}{\partial \boldsymbol{\phi}_i} \ln p(z_n = k \mid \mathbf{x}_n, \boldsymbol{\Phi}) \tag{19}$$

1.4

$$\frac{\partial}{\partial \boldsymbol{\theta}_i} \ln p(\mathbf{y} \mid \mathbf{X}, \boldsymbol{\Theta}, \boldsymbol{\Phi}) = \sum_{n=1}^{N} r_{ni} \frac{\partial}{\partial \boldsymbol{\theta}_i} \ln p(y_n \mid \mathbf{x}_n, \boldsymbol{\theta}_i, z_n = i) \tag{20}$$

$$= \sum_{n=1}^{N} r_{ni} \frac{\partial}{\partial \boldsymbol{\theta}_i} \ln \Big[\exp(\boldsymbol{\theta}_i^\top \mathbf{x}_n) \exp(-\exp(\boldsymbol{\theta}_i^\top \mathbf{x}_n) y_n) \Big] \tag{21}$$

$$= \sum_{n=1}^{N} r_{ni} \frac{\partial}{\partial \boldsymbol{\theta}_i} \Big[\boldsymbol{\theta}_i^\top \mathbf{x}_n - \exp(\boldsymbol{\theta}_i^\top \mathbf{x}_n) y_n \Big] \tag{22}$$

$$= \sum_{n=1}^{N} r_{ni} \Big[\mathbf{x}_n^\top - y_n \frac{\partial}{\partial \boldsymbol{\theta}_i} \exp(\boldsymbol{\theta}_i^\top \mathbf{x}_n) \Big] \tag{23}$$

$$= \sum_{n=1}^{N} r_{ni} \Big[\mathbf{x}_n^\top - y_n \exp(\boldsymbol{\theta}_i^\top \mathbf{x}_n) \mathbf{x}_n^\top \Big] \tag{24}$$

$$= \sum_{n=1}^{N} r_{ni} \Big[1 - y_n \exp(\boldsymbol{\theta}_i^\top \mathbf{x}_n) \Big] \mathbf{x}_n^\top \tag{25}$$

$\frac{\partial}{\partial \boldsymbol{\phi}_i} \ln p(\mathbf{y} \mid \mathbf{X}, \boldsymbol{\Theta}, \boldsymbol{\Phi})$: next page.

$$\frac{\partial}{\partial \boldsymbol{\phi}_i} \ln p(\mathbf{y} \mid \mathbf{X}, \boldsymbol{\Theta}, \boldsymbol{\Phi}) = \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} \frac{\partial}{\partial \boldsymbol{\phi}_i} \ln p(z_n = k \mid \mathbf{x}_n, \boldsymbol{\Phi}) \tag{26}$$

$$= \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} \frac{\partial}{\partial \boldsymbol{\phi}_i} \ln \frac{\exp(\boldsymbol{\phi}_k^\top \mathbf{x}_n)}{\sum_j \exp(\boldsymbol{\phi}_j^\top \mathbf{x}_n)} \tag{27}$$

$$= \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} \frac{\partial}{\partial \boldsymbol{\phi}_i} \Big[\boldsymbol{\phi}_k^\top \mathbf{x}_n - \ln \sum_j \exp(\boldsymbol{\phi}_j^\top \mathbf{x}_n) \Big] \tag{28}$$

$$= \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} \Big[\mathbf{x}^\top \delta_{ik} - \frac{\frac{\partial}{\partial \boldsymbol{\phi}_i} \sum_j \exp(\boldsymbol{\phi}_j^\top \mathbf{x}_n)}{\sum_j \exp(\boldsymbol{\phi}_j^\top \mathbf{x}_n)} \Big] \tag{29}$$

$$= \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} \Big[\mathbf{x}^\top \delta_{ik} - \frac{\exp(\boldsymbol{\phi}_i^\top \mathbf{x}_n) \mathbf{x}_n^\top}{\sum_j \exp(\boldsymbol{\phi}_j^\top \mathbf{x}_n)} \Big] \tag{30}$$

$$= \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} \Big[\mathbf{x}^\top \delta_{ik} - \pi_{ni} \mathbf{x}_n^\top \Big] \tag{31}$$

$$= \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} \Big[\delta_{ik} - \pi_{ni} \Big] \mathbf{x}^\top \tag{32}$$

$$= \sum_{n=1}^{N} \Big[r_{ni} - \pi_{ni} \Big] \mathbf{x}^\top \tag{33}$$

2 Quadratic Discriminant Analysis (QDA)

2.1

$$p(\mathbf{x}_n, \mathcal{C}_k) = p(\mathcal{C}_k) p(\mathbf{x}_n \mid \mathcal{C}_k) \tag{34}$$

$$= \pi_k \mathcal{N}(\mathbf{x}_n \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \tag{35}$$

2.2

Here we use $\boldsymbol{\theta}$ to denote $\pi_1, \ldots, \pi_K, \boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_K, \boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_K$.

$$p(\mathbf{T}, \mathbf{X} \mid \boldsymbol{\theta}) = \prod_{n=1}^{N} p(\mathbf{t}_n, \mathbf{x}_n \mid \boldsymbol{\theta}) \tag{36}$$

$$= \prod_{n=1}^{N} p(\mathbf{t}_n \mid \boldsymbol{\theta}) p(\mathbf{x}_n \mid \mathbf{t}_n, \boldsymbol{\theta}) \tag{37}$$

$$= \prod_{n=1}^{N} \prod_{k=1}^{K} \Big[p(t_{nk} \mid \boldsymbol{\theta}) p(\mathbf{x}_n \mid t_{nk}, \boldsymbol{\theta}) \Big]^{t_{nk}} \tag{38}$$

$$= \prod_{n=1}^{N} \prod_{k=1}^{K} \Big[\pi_k p(\mathbf{x}_n \mid \mathcal{C}_k) \Big]^{t_{nk}} \tag{39}$$

$$= \prod_{n=1}^{N} \prod_{k=1}^{K} \Big[\pi_k \mathcal{N}(\mathbf{x}_n \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \Big]^{t_{nk}} \tag{40}$$

$$\ln p(\mathbf{T}, \mathbf{X} \mid \boldsymbol{\theta}) = \sum_{n=1}^{N} \sum_{k=1}^{K} \ln \Big[\pi_k p(\mathbf{x}_n \mid \mathcal{C}_k) \Big]^{t_{nk}} \tag{41}$$

$$= \sum_{n=1}^{N} \sum_{k=1}^{K} t_{nk} \ln \Big[\pi_k p(\mathbf{x}_n \mid \mathcal{C}_k) \Big] \tag{42}$$

$$= \sum_{n=1}^{N} \sum_{k=1}^{K} t_{nk} \ln \Big[\pi_k \mathcal{N}(\mathbf{x}_n \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \Big] \tag{43}$$

2.3

Lagrangian:

$$L = \ln p(\mathbf{T}, \mathbf{X} \mid \boldsymbol{\theta}) + \lambda(1 - \sum_{k=1}^{K} \pi_k) \tag{44}$$

$$= \sum_{n=1}^{N} \sum_{k=1}^{K} t_{nk} \ln \Big[\pi_k \mathcal{N}(\mathbf{x}_n \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)\Big] + \lambda(1 - \sum_{k=1}^{K} \pi_k) \tag{45}$$

2.4

$$\frac{\partial L}{\partial \pi_k} = \sum_{n=1}^{N} \frac{\partial}{\partial \pi_k} \sum_{k'=1}^{K} t_{nk'} \ln \Big[\pi_{k'} \mathcal{N}(\mathbf{x}_n \mid \boldsymbol{\mu}_{k'}, \boldsymbol{\Sigma}_{k'})\Big] + \lambda \frac{\partial}{\partial \pi_k}(1 - \sum_{k'}^{K} \pi_{k'}) \tag{46}$$

$$= \sum_{n=1}^{N} t_{nk} \frac{\partial}{\partial \pi_k} \ln \Big[\pi_k \mathcal{N}(\mathbf{x}_n \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)\Big] - \lambda \frac{\partial}{\partial \pi_k} \sum_{k'}^{K} \pi_{k'} \tag{47}$$

$$= \sum_{n=1}^{N} t_{nk} \frac{1}{\pi_k} - \lambda \tag{48}$$

$$= \frac{N_k}{\pi_k} - \lambda, \qquad N_k = \sum_{n=1}^{N} t_{nk} \tag{49}$$

$$\stackrel{!}{=} 0 \tag{50}$$

$$\pi_k = \frac{N_k}{\lambda} \tag{51}$$

$$\tag{52}$$

Summing over all k gives us λ:

$$\sum_{k=1}^{K} \pi_k = 1 = \frac{\sum_{k=1}^{K} N_k}{\lambda} \tag{53}$$

$$\lambda = \sum_{k=1}^{K} N_k = N \tag{54}$$

So for π_{ML} we get:

$$\pi_k = \frac{N_k}{N} \tag{55}$$

2.5

In the following we use D to denote the size of vector $\mathbf{x}_n$.

$$\frac{\partial L}{\partial \boldsymbol{\mu}_k} = \sum_{n=1}^{N} \frac{\partial}{\partial \boldsymbol{\mu}_k} \sum_{k'=1}^{K} t_{nk'} \ln \Big[\pi_{k'} \mathcal{N}(\mathbf{x}_n \mid \boldsymbol{\mu}_{k'}, \boldsymbol{\Sigma}_{k'}) \Big] \tag{56}$$

$$= \sum_{n=1}^{N} t_{nk} \frac{\partial}{\partial \boldsymbol{\mu}_k} \ln \Big[\pi_k \mathcal{N}(\mathbf{x}_n \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \Big] \tag{57}$$

$$= \sum_{n=1}^{N} t_{nk} \frac{\partial}{\partial \boldsymbol{\mu}_k} \ln \mathcal{N}(\mathbf{x}_n \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \tag{58}$$

$$= \sum_{n=1}^{N} t_{nk} \frac{\partial}{\partial \boldsymbol{\mu}_k} \ln \left(\frac{1}{(2\pi)^{D/2}} \frac{1}{|\boldsymbol{\Sigma}_k|^{1/2}} \exp \Big[-\frac{1}{2} (\mathbf{x}_n - \boldsymbol{\mu}_k)^\top \boldsymbol{\Sigma}_k^{-1} (\mathbf{x}_n - \boldsymbol{\mu}_k) \Big] \right) \tag{59}$$

$$= \sum_{n=1}^{N} t_{nk} \frac{\partial}{\partial \boldsymbol{\mu}_k} \ln \exp \Big[-\frac{1}{2} (\mathbf{x}_n - \boldsymbol{\mu}_k)^\top \boldsymbol{\Sigma}_k^{-1} (\mathbf{x}_n - \boldsymbol{\mu}_k) \Big] \tag{60}$$

$$= \sum_{n=1}^{N} -\frac{t_{nk}}{2} \frac{\partial}{\partial \boldsymbol{\mu}_k} (\mathbf{x}_n - \boldsymbol{\mu}_k)^\top \boldsymbol{\Sigma}_k^{-1} (\mathbf{x}_n - \boldsymbol{\mu}_k) \tag{61}$$

$$= \sum_{n=1}^{N} -\frac{t_{nk}}{2} \big[-2 (\mathbf{x}_n - \boldsymbol{\mu}_k)^\top \boldsymbol{\Sigma}_k \big] \tag{62}$$

$$= \sum_{n=1}^{N} t_{nk} (\mathbf{x}_n - \boldsymbol{\mu}_k)^\top \boldsymbol{\Sigma}_k \tag{63}$$

$$= \sum_{n=1}^{N} t_{nk} \mathbf{x}_n^\top \boldsymbol{\Sigma}_k - \sum_{n=1}^{N} t_{nk} \boldsymbol{\mu}_k^\top \boldsymbol{\Sigma}_k \tag{64}$$

$$= \sum_{n \in \mathcal{C}_k}^{N} \mathbf{x}_n^\top \boldsymbol{\Sigma}_k - N_k \boldsymbol{\mu}_k^\top \boldsymbol{\Sigma}_k \tag{65}$$

$$\stackrel{!}{=} 0 \tag{66}$$

$$N_k \boldsymbol{\mu}_k^\top \boldsymbol{\Sigma}_k = \sum_{n \in \mathcal{C}_k}^{N} \mathbf{x}_n^\top \boldsymbol{\Sigma}_k \tag{67}$$

We can take the transpose on each side, and eliminate $\boldsymbol{\Sigma}_k$ from the equation. So for $\boldsymbol{\mu}_{\text{ML}}$ we get:

$$\boldsymbol{\mu}_k = \frac{1}{N_k} \sum_{n \in \mathcal{C}_k}^{N} \mathbf{x}_n \tag{68}$$

2.6

Using Eq. 2.122 in Bishop (p.94) we find for $\boldsymbol{\Sigma}_{\text{ML}}$:

$$\boldsymbol{\Sigma}_k = \frac{1}{N_k} \sum_{n \in \mathcal{C}_k}^{N} (\mathbf{x}_n - \boldsymbol{\mu}_k)(\mathbf{x}_n - \boldsymbol{\mu}_k)^\top, \tag{69}$$

where $\boldsymbol{\mu}_k$ is the ML-estimator for the k'th class mean. Since each class has its own covariance matrix, we made Eq. 2.122 class-specific.

2.7

So we have:

$$\pi_k = \frac{N_k}{N} \tag{70}$$

$$\boldsymbol{\mu}_k = \frac{1}{N_k} \sum_{n \in \mathcal{C}_k}^{N} \mathbf{x}_n \tag{71}$$

$$\boldsymbol{\Sigma}_k = \frac{1}{N_k} \sum_{n \in \mathcal{C}_k}^{N} (\mathbf{x}_n - \boldsymbol{\mu}_k)(\mathbf{x}_n - \boldsymbol{\mu}_k)^\top \tag{72}$$

π_k tells us what the probability is of class k, and the ML-estimator (Eq. 70) tells us that that probability can be estimated by taking the number of data points that belong to class k, and dividing it by the total number of data points N. The second one is also very intuitive. $\boldsymbol{\mu}_k$ tells us what the mean is of the data points belonging to class k, and The ML-estimator (Eq. 71) shows that that mean can be estimated by summing up all data points belonging to class k, and dividing that by the number of data points belonging to class k. Lastly, $\boldsymbol{\Sigma}_k$ tells us how the points in class k are correlated, and the ML-estimator (Eq. 72) shows that these correlations can be estimated by comparing each data point in that class with the mean of that class (and dividing the sum by the number of points in that class, N_k).

3 Principal Component Analysis

3.1

$$z_{ni} = \mathbf{u}_i^\top \mathbf{x}_n \tag{73}$$

3.2

Since our data set has zero mean for each dimension, the empirical mean of the projection z_i across all points $\mathbf{x}_n$ is zero.

$$\bar{z}_i = \frac{1}{N} \sum_{n=1}^{N} z_{ni} = \mathbf{u}_i^\top \bar{\mathbf{x}} = 0 \tag{74}$$

3.3

$$\mathbb{V}(z_i) = \frac{1}{N} \sum_{n=1}^{N} (\mathbf{u}_i^\top \mathbf{x}_n - \mathbf{u}_i^\top \bar{\mathbf{x}})^2 = \mathbf{u}_i^\top \mathbf{S} \mathbf{u}_i \tag{75}$$

3.4

$$\mathbb{V}(z_i) = \mathbf{u}_i^\top \mathbf{U} \boldsymbol{\Lambda} \mathbf{U}^\top \mathbf{u}_i \tag{76}$$

Since U is an orthonormal matrix, we get $\mathbf{u}_i^\top U = \mathbf{e}_i^\top$, where $\mathbf{e}_i$ is a vector of zeros, with the i'th element being equal to 1. Also, because Λ is a diagonal matrix consisting of eigenvalues λ_i of $\mathbf{S}$, we get

$$\mathbb{V}(z_i) = \mathbf{e}_i^\top \boldsymbol{\Lambda} \mathbf{e}_i \tag{77}$$

$$= \lambda_i \tag{78}$$

3.5

K should be chosen such that the following holds:

$$\frac{\sum_{k=1}^{K} \lambda_k}{\sum_{d=1}^{D} \lambda_d} \geq 0.99 \tag{79}$$

This condition is most easily met if the eigenvalues are sorted in descending order.