1. In Principal Components Analysis, given a y and its a low-dimensional projection x with the following equation: $\mathrm{y} = \mathrm{w}^{\mathrm{T}} x$, please answer the following questions:
   (a) Please explain the connection between w and y (Use math). [5pt]
   (b) Assume the covariance matrix of y is $\begin{matrix} 16 & 0 & 2 \\ 0 & 25 & 0 \\ 2 & 0 & 4 \end{matrix}$ and we set $\mathrm{k} = 1$ in PCA, please determine the matric w. [10pt]
2. Define a multivariate Bernoulli mixture where inputs are binary and derive the EM equation.[10pt]
3. Generalize the Gini index and the misclassification error for K > 2 classes. Generalize misclassification error to risk, taking a loss function into account.[5pt]
4. Show that the derivative of the softmax, $\mathrm{y_i} = \frac{\exp(\mathrm{a_i})}{\sum_j \exp(a_j)}$, is $\frac{\partial y_i}{\partial a_j} = y_i(\delta_{ij} - y_j)$ where δ_{ij} is 1 if $\mathrm{i} = \mathrm{j}$ and 0 otherwise. [10pt]

1. ANS

(a)

- The projection of $\boldsymbol{x}$ on the direction of $\boldsymbol{w}$ is: $z = \boldsymbol{w}^T\boldsymbol{x}$
- Find $\boldsymbol{w}$ such that $\boxed{\mathrm{Var}(z)}$ is maximized

information

$$\begin{aligned}\mathrm{Var}(z) &= \mathrm{Var}(\boldsymbol{w}^T\boldsymbol{x}) = E[(\boldsymbol{w}^T\boldsymbol{x} - \boldsymbol{w}^T\boldsymbol{\mu})^2] \\ &= E[(\boldsymbol{w}^T\boldsymbol{x} - \boldsymbol{w}^T\boldsymbol{\mu})(\boldsymbol{w}^T\boldsymbol{x} - \boldsymbol{w}^T\boldsymbol{\mu})] \\ &= E[\boldsymbol{w}^T(\boldsymbol{x} - \boldsymbol{\mu})(\boldsymbol{x} - \boldsymbol{\mu})^T\boldsymbol{w}] \\ &= \boldsymbol{w}^T E[(\boldsymbol{x} - \boldsymbol{\mu})(\boldsymbol{x} - \boldsymbol{\mu})^T]\boldsymbol{w} = \boldsymbol{w}^T \Sigma \boldsymbol{w}\end{aligned}$$

where $\mathrm{Var}(\boldsymbol{x}) = E[(\boldsymbol{x} - \boldsymbol{\mu})(\boldsymbol{x} - \boldsymbol{\mu})^T] = \Sigma$

(b)

- First projection: Maximize $\mathrm{Var}(z_1)$ subject to $\|\boldsymbol{w}_1\|=1$

$$\max_{\mathbf{w}_1} \mathbf{w}_1^T \Sigma \mathbf{w}_1 - \alpha\left(\mathbf{w}_1^T\mathbf{w}_1 - 1\right)$$

A Lagrange problem

$\boxed{\Sigma \boldsymbol{w}_1 = \alpha \boldsymbol{w}_1}$ that is, $\boldsymbol{w}_1$ is an eigenvector of Σ

Choose the one with the largest eigenvalue for $\mathrm{Var}(z)$ to be max

For derive PCA with k=1, you need to compute the biggest eigenvalue and the corresponding eigenvector for covariance matrix of y.

The equation of finding eigenvalues could be written as:

$\det\begin{pmatrix}16 & 0 & 2 \\ 0 & 25 & 0 - \lambda\mathbf{I} \\ 2 & 0 & 4\end{pmatrix} = 0$, we could get three eigenvalues

$\lambda = 25, 16.3245, 3.6754$

the biggest one is 25, and its corresponding eigenvector is [0,1,0]=W

(Remind: This means that the PCA only keep the second feature of X)

2. ANS

When the components are multivariate Bernouilli, we have binary vectors that are d-dimensional. Assuming that the dimensions are independent, we have (see section 5.7)

$$p_i(\mathbf{x}^t|\Phi) = \prod_{j=1}^{d} p_{ij}^{x_j^t}(1-p_{ij})^{1-x_j^t}$$

where $\Phi^l = \{p_{i1}^l, p_{i2}^l, \ldots, p_{id}^l\}_{i=1}^k$. The E-step does not change (equation 7.9). In the M-step, for the component parameters $p_{ij}, i = 1, \ldots, k, j = 1, \ldots, d$, we maximize

$$\begin{aligned} \mathcal{Q}' &= \sum_t \sum_i h_i^t \log p_i(\mathbf{x}^t|\phi^l) \\ &= \sum_t \sum_i h_i^t \sum_j x_j^t \log p_{ij}^l + (1-x_j^t)\log(1-p_{ij}^l) \end{aligned}$$

Taking the derivative with respect to p_{ij} and setting it equal to 0, we get

$$p_{ij}^{l+1} = \frac{\sum_t h_i^t x_j^t}{\sum_t h_j^t}$$

Note that this is the same as in equation 5.31, except that estimated "soft" labels h_i^t replace the supervised labels r_i^t.

3. ANS

- Gini index with $K > 2$ classes: $\phi(p_1, p_2, \ldots, p_K) = \sum_{i=1}^K \sum_{j<i} p_i p_j$
- Misclassification error: $\phi(p_1, p_2, \ldots, p_K) = 1 - \max_{i=1}^K p_i$
- Risk: $\phi_\Lambda(p_1, p_2, \ldots, p_K) = \min_{i=1}^K \sum_{k=1}^K \lambda_{ik} p_k$ where Λ is the $K \times K$ loss matrix (equation 3.7).

4. ANS

Given that

$$y_i = \frac{\exp a_i}{\sum_j \exp a_j}$$

for $i = j$, we have

$$\begin{aligned}
\frac{\partial y_i}{\partial a_i} &= \frac{\exp a_i \left(\sum_j \exp a_j\right) - \exp a_i \exp a_j}{\left(\sum_j \exp a_j\right)^2} \\
&= \frac{\exp a_i}{\sum_j \exp a_j} \left(\frac{\sum_j \exp a_j - \exp a_i}{\sum_j \exp a_j}\right) \\
&= y_i(1 - y_i)
\end{aligned}$$

and for $i \neq j$, we have

$$\begin{aligned}
\frac{\partial y_i}{\partial a_j} &= \frac{-\exp a_i \exp a_j}{\left(\sum_j \exp a_j\right)^2} \\
&= -\left(\frac{\exp a_i}{\sum_j \exp a_j}\right) \left(\frac{\sum_j \exp a_j}{\sum_j \exp a_j}\right) \\
&= y_i(0 - y_j)
\end{aligned}$$

which we can combine in one equation as

$$\frac{\partial y_i}{\partial a_j} = y_i(\delta_{ij} - y_j)$$