4.11 For clarity's sake, we can visualise $\boldsymbol{\phi}$ as:

$$\boldsymbol{\phi} = \left[[\phi_1]^T \quad [\phi_2]^T \quad \ldots \quad [\phi_m]^T \ldots \quad [\phi_M]^T\right]^T$$

where each $[\phi_m]^T$ is an L sized vector, looking something like $[\quad 0 \quad \ldots \quad 1 \quad \ldots 0 \quad]$, with one 1 and (L-1) 0s since it is a 1-of-L binary coding scheme.

We can use ϕ_{ml} to represent the coding scheme such that $\phi_{ml} = 0$ or 1.

The class conditional distributions are of the form:

$$p(\boldsymbol{\phi}|\mathcal{C}_k) = \prod_{m=1}^{M} \prod_{l=1}^{L} \mu_{kml}^{\phi_{ml}}$$

where μ_{kml} is the probability of $\phi_{ml} = 1$ given class k.

The likelihood function is then given by:

$$p(\boldsymbol{\Phi}| \ldots \mu_{ml} \ldots) = \prod_{i=1}^{N} \prod_{m=1}^{M} \prod_{l=1}^{L} \mu_{kml}^{\phi_{iml}}$$

Substituting into (4.63) then gives:

$$a_k(\boldsymbol{\phi}) = \sum_{m=1}^{M} \sum_{l=1}^{L} \{\phi_{ml} \ln \mu_{kml}\} + \ln p(\mathcal{C}_k)$$

which is a linear function of the components of $\boldsymbol{\phi}$.