Pseudo Code :- Mixture of experts

Init ∂ik

Repeat

E: calc ∂ik : $\dfrac{P_{ik} * P(\text{output } i / \text{class } k)}{\sum_{\forall k} P_{ik} * P(\text{output } i / \text{class } k)}$

M step:-

Step 1 Multiclass logistic regression using ∂_{ik} as labels

init W ~~weight~~ vector (L_2 Regularized)

@ using numpy for softmax calulatn.

$$\text{yhat} = \frac{\exp(a_K)}{\sum_j \exp(a_j)} \qquad a_K = \cancel{\phi \cdot w}\ \cancel{x}\ \phi W^T.$$

numpy

~~Er~~ ~~Err~~ ~~Cr~~ Crossentropy

$$= -\sum_{n=1}^{N} \sum_{k=1}^{K} t_{nk} \ln y_{nk} + \frac{\lambda}{2} \sum_{\forall j} \|w_j\|^2.$$

grad1 : $\cancel{\sum_{n=1}^{N}} + \phi^T(\hat{y}\text{hat} - y) + \frac{\lambda}{2} W^T$

grad2 : $-\sum_{n=1}^{N} y_{nk}(I_{kj} - y_{nj}) \phi_n \phi_n^T + \lambda I$

Using minimize functionality in python to minize the Crossentropy using Newton-CG Method & both grad1 & grad2 are passed into minimize.

Step2 Running k times di L_2-logistic Regression (scikt-learn) where example weights as ∂_{ik} & all the parameter kept at ~~default~~ default Values.

~~Runnig~~ Running the process untill $\partial|\partial_{new} - \partial_{odd}| < 0.001$