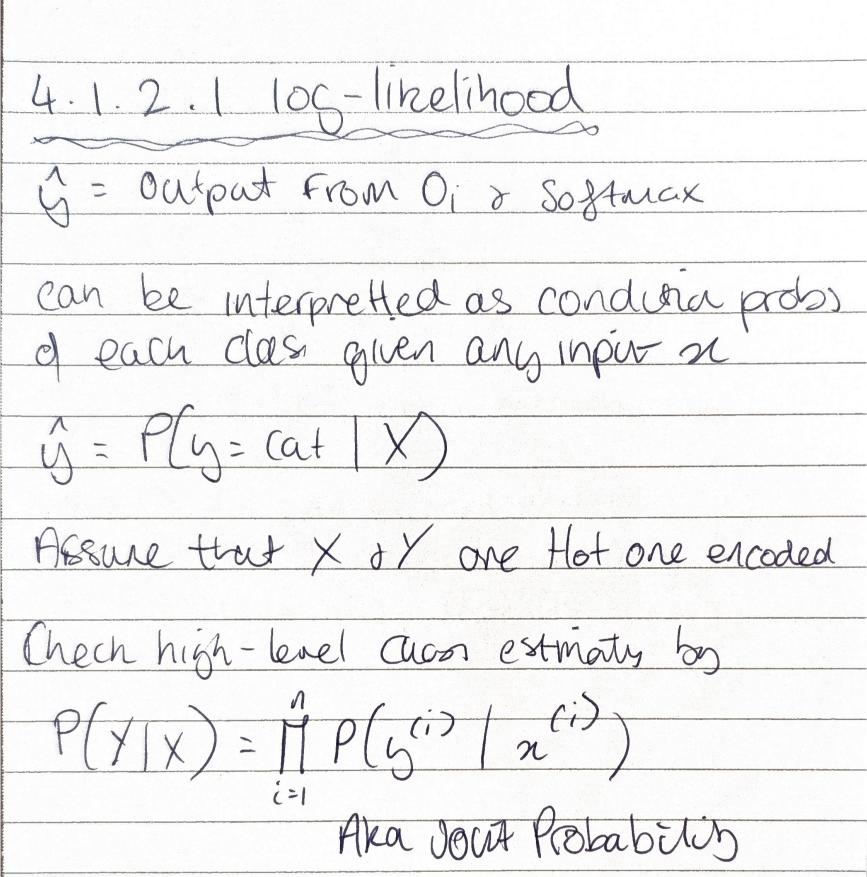


4.1.1.2 Softmax
for probs output to be valid the new to be 70 & sum to 1
Cant treat Claff this regression as no garentee these rules hold
We need to "Squise" model output
best way = expo funct P(y=1) or ex. O
Result: - Cond prob increase w/Oi - Monotonic
- All probs are non-neg - Sun Probs sum to I
This process is called Normalization
Putting all the together we get the soptilax function

 $\hat{y} = Softm(o)$ where $\hat{y}_i = exp(o_i)$ $= \frac{1}{2}, exp(o_i)$ 0 = Class output model Softmax Aguoid Func for >2 Classes arguax vi = Arguax O; 4.1.2 loss function Now that we are able to map features x to probabilities is Need a way to opt accurrency of map Rely on max likelihood Funcher



Maximingring the prod of ferms is awkward So we take the wes of los of ferm -lg(P(YIX))= \(\int - los P(G') (n') \)

this gives way to the loss function $\Sigma L(g^{(i)}, \hat{g}^{(i)})$ where $l(y, \hat{y}) = -\sum_{i=1}^{q} y_i \cdot log \hat{y}_i$ this is called cross-entrophy loss recally = Hot one encoded vector = the 1 traggers the corle like a Ewith, the Os turnon