

This handout includes space for every question that requires a written response. Please feel free to use it to handwrite your solutions (legibly, please). If you choose to typeset your solutions, the `README.md` for this assignment includes instructions to regenerate this handout with your typeset L<sup>A</sup>T<sub>E</sub>X solutions.

---

1.a

1.b

1.c

The log-likelihood of an example  $(x^{(i)}, y^{(i)})$  is defined as  $\ell(\theta) = \log p(y^{(i)} | x^{(i)}; \theta)$ . To derive the stochastic gradient ascent rule, use the results in part (a) and the standard GLM assumption that  $\eta = \theta^T x$ .

$$\begin{aligned} \frac{\partial \ell(\theta)}{\partial \theta_j} &= \frac{\partial \log p(y^{(i)} | x^{(i)}; \theta)}{\partial \theta_j} \\ &= \frac{\partial \log \left( \frac{1}{y^{(i)}!} \exp(\eta^T y^{(i)} - e^\eta) \right)}{\partial \theta_j} \\ &= \end{aligned}$$

Thus the stochastic gradient ascent update rule should be:

$$\theta_j := \theta_j + \alpha \frac{\partial \ell(\theta)}{\partial \theta_j},$$

which reduces here to:

2.a

2.b

2.c

2.d

3.ai

3.ii

3.iii

3.c