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 LaTeX solutions.

1.a

$$\begin{aligned}
p(y;\lambda) &= \frac{exp(-\lambda)\lambda^y}{y!} \\
&= b(y)exp(-\lambda)\lambda^y \text{ where } b(y) = \frac{1}{y!} \\
&= b(y)exp(-\lambda + log(\lambda)y) \\
&= b(y)exp(\eta y - exp(\eta)) \text{ where } \eta = log(\lambda) \\
& b(y)exp(\eta^t y - \alpha(\eta)) \text{ where } \alpha(\eta) = exp(\eta)
\end{aligned}$$

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

We will show that $K = K_1 + K_2$ is symmetric and PSD.

First symmetric. This will follow since K_1 and K_2 are symmetric.
Denote the (i, j) entry of K, as $K^{(i,j)}$. Then

$$\begin{aligned} K^{(i,j)} &= K_1^{(i,j)} + K_2^{(i,j)} \\ &= K_1^{(j,i)} + K_2^{(j,i)} \text{ since } K_1 \text{ and } K_2 \text{ are symmetric} \\ &= K^{(j,i)} \end{aligned}$$

Now for PSD. Since K_1 and K_2 are PSD then
$\forall v$ we know $v^T K_1 v \geq 0$ and $v^T K_2 v \geq 0$
Therefore $v^T K v = v^T (K_1 + K_2) v = v^T K_1 v + v^T K_2 v \geq 0$

2.b

Let $K_2 = -2K_1$ for some $K_1 \neq 0$.
Then for some v where $v^T K_1 v = c > 0$
$v^T K v = v^T (K_1 + K_2) v = v^T K_1 v - 2v^T K_1 v = c - 2c = -c < 0$
So $K = K_1 - K_2$ is not a kernel.

2.c

We will show that $K = aK_1$ is symmetric and PSD
given that K_1 is symmetric and PSD and $a \in \mathbb{R}^+$ i.e. $a > 0$.

First symmetric. This will follow since K_1 is symmetric.
Denote the (i,j) entry of K, as $K^{(i,j)}$. Then

$$\begin{aligned} K^{(i,j)} &= aK_1^{(i,j)} \\ &= aK_1^{(j,i)} \text{ since } K_1 \text{ is symmetric} \\ &= aK^{(j,i)} \end{aligned}$$

Now for PSD. Since K_1 is PSD then
$\forall v$ we know $v^T K_1 v \geq 0$
Therefore $v^T K v = v^T (aK_1) v = a v^T K_1 v = ac$ where $a > 0$ and $c >= 0$ so $ac \geq 0$

2.d

Choose $K_1 \neq 0$ and for some v where $v^T K_1 v > 0$ then
$v^T K v = v^T(-aK_1)v = -a(v^T K_1 v) = -ac < 0$

So $K = -aK_1$ is not a kernel.

3.ai

Represent θ as a linear combination of data points we have seen so far, as in

$$\theta = \sum_{i=1}^{m} \beta_i \phi(x^{(i)})$$

Thus we only need the m β values

3.aii

$$\begin{aligned} h_{\theta^{(i)}}(x^{(i+1)}) &= g({\theta^{(i)}}^T \phi(x^{(i+1)})) \\ &= g(\sum_{i=1}^{n} \beta_i \phi(x^{(i)}) \phi(x^{(i+1)})) \\ &= g(\sum_{i=1}^{n} \beta_i K(x^{(i)}, x)) \end{aligned}$$

Where K is the kernel function.

3.aiii

Since we are representing θ via the β values we need a rule to update the β values.

$$\beta_j^{(i+1)} := \beta_j^{(i)} + \alpha(y^{(i+1)} - g(\sum_{i=1}^{n} \beta_i^{(i)} K(x^{(i)}, x^{(i+1)}))$$

3.c