```latex
% This is a template for doing homework assignments in LaTeX

\documentclass{article} % This command is used to set the type of document you are working on such as an article, book, or presenation

\usepackage{geometry} % This package allows the editing of the page layout
\usepackage{amsmath}  % This package allows the use of a large range of mathematical formula, commands, and symbols
\usepackage{graphicx}  % This package allows the importing of images
\usepackage[thinc]{esdiff} % This package allows for use of helpful diff functions
\usepackage{cancel} % for the \cancel command
\usepackage{hyperref} % add links

\newcommand{\question}[2][]{\begin{flushleft}
      \textbf{Question #1}: \textit{#2}

\end{flushleft}}
\newcommand{\sol}{\textbf{Solution}:} %Use if you want a boldface solution line
\newcommand{\maketitletwo}[2][]{\begin{center}
      \Large{\textbf{Assignment #1}

         CS 229 2018 Fall} % Name of course here
      \vspace{5pt}

      \normalsize{Eric Lagnese  % Your name here

         \today}       % Change to due date if preferred
      \vspace{15pt}

\end{center}}
\begin{document}
   \maketitletwo[1]  % Optional argument is assignment number
   %Keep a blank space between maketitletwo and \question[1]

   \question[1]{Find the Hessian:
      \begin{align*}
         J(\theta)=
         -\frac{1}{m}\sum_{i=1}^{m}y^{(i)}log(h_\theta(x^{(i)}))+(1-y^{(i)})log(
         1-h_\theta(x^{(i)})) \\
      \end{align*}
   }

   Notes: To me the most helpful thing in solving this problem was using mathematicalmonk for reference.  Specifically lessons 15.3-15.6.  If you get stuck this is an excellent resource however I will go through step-by-step on how I solved this problem and possible explain some things that weren't trivial to you before.
   \paragraph{Background} The Hessian describes the curviture and rate of change of a surface.  It is found by find the gradient (the first set of partial derivatives), then taking the partial derivatives of the gradient. You can consider this similar to taking the second derivative of a function to find the maximums and minimums.  The significance of the Hessian lies in its ability to help us locate the minimum error by analyzing the error function.
   \paragraph{Solution}
   Lets move from this summation to some value for i. It will just look easier notation wise and helps my brain.  Consider that each operation will apply to every iteration of i.
```

will apply to every iteration of i.
The equation will look like:
\begin{align*}
J(\theta)= -\frac{1}{m}y\cdot
log(\frac{1}{1+e^{-\theta^Tx}})+(1-y)log(1-\frac{1}{1+e^{-\theta^Tx}})
\\
\end{align*}

First I'm going to use some shorthand.
$\alpha = \frac{1}{1+e^{-\theta^Tx}}$
So we have:
\begin{align*}
J(\theta)= -\frac{1}{m}y\cdot log(\alpha)+(1-y)log(1-\alpha)
\end{align*}
Now we ultimately need to take two partial derivatives. One with respect to θ_j and the other with respect to θ_k.

\begin{align*}
\diffp{J(\theta)}{{\theta_j}} = \diffp{}{{\theta_j}} \left
[-\frac{1}{m}y\cdot log(\alpha)+(1-y)log(1-\alpha) \right]
\end{align*}

First lets break it up and deal with $ y^{(i)}log(\alpha)$ first. This is simple if you use the home work hint:
$\alpha'=\alpha(1-\alpha)(\theta^T x)'$

Left-hand side of the plus first:
\begin{align*}
\diffp{}{{\theta_j}} y\cdot log(\alpha) = y\cdot \frac{1}{\alpha}\alpha'
\\
y \frac{1}{\alpha}\alpha' = y\cdot
\frac{1}{\alpha}\alpha(1-\alpha)(\theta^T x)'
\end{align*}

If we look at $(\theta^T x)'$, the best way I thought about its simplification is if you expand it and then try to take the partial derivative. So $(\theta^T x)'$ becomes $\theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + ... + \theta_n x_n$. Now if you were to take the derivative with respect to θ_1 for example, most of the items would go to zero because they are constants in that dimension. So your result would be x_1. Now do the same but for θ_j and that's how you get $(\theta^T x)'=x_j$.
\begin{align*}
y \frac{1}{\alpha}\alpha' = y \frac{1}{\alpha}\alpha(1-\alpha)(x_j)
\end{align*}
Reorganizing:
\begin{align*}
= y \frac{x_j\alpha(1-\alpha)}{\alpha} \\
= y\cdot x_j(1-\alpha)
\end{align*}
That should be simplified enough for us to move on. Now lets do the same for the right-hand side of the plus: $(1-y)log(1-\alpha)$.
\begin{align*}
= \diffp {} {{\theta_j}} (1-y)log(1-\alpha)\\
= (1-y) \frac{1}{1-\alpha}(- \alpha')
\end{align*}
Again using the substitution of the hint
($\alpha'=\alpha(1-\alpha)(\theta^T x)'$):
\begin{align*}

```latex
      = (1-y) \frac{1}{1-\alpha}(-\alpha(1-\alpha)(\theta^T x)')\\
      = (-1+y) \frac{\alpha(1-\alpha)}{1-\alpha}(\theta^T x)'\\
      = (-1+y) \frac{\alpha\cancel{(1-\alpha)}}{\cancel{(1-\alpha)}}(x_j) \\
      = (-1+y) \alpha(x_j)
   \end{align*}
   Putting it all together we have:
   \begin{align*}
      = - \frac{1}{m} [y\cdot x_j(1-\alpha) + (-1 + y) \alpha(x_j) ] \\
      = - \frac{1}{m} [(y-y \alpha) + (-\alpha+y\cdot\alpha) ]x_j \\
      = - \frac{1}{m} [(y - \alpha)x_j ] \\
   \end{align*}
      substituting  we get
\begin{align*}
      = - \frac{1}{m} \sum_{i=1}^{m} (y^{(i)}  - h_\theta(x^{(i)}))x_j ^{(i)}
   \end{align*}

   Now we to apply $\diffp{}{{\theta_k}}$ to this equation.  I will start with
   the simplified equation using $\alpha$'s.
   \begin{align*}
      = \diffp{}{{\theta_k}}\left[-\frac{1}{m} [(y - \alpha)x_j ] \right]
   \end{align*}

   Take the derivative one at a time.
   \[
   \begin{minipage}{.5\linewidth}
      \begin{align*}
         = \diffp{}{{\theta_k}}yx_j  \\
         =0
      \end{align*}
   \end{minipage}%
   \begin{minipage}{.5\linewidth}
      \begin{align*}
         = \diffp{}{{\theta_k}} -\alpha x_j  \\
         =  -\alpha(1-\alpha)(\theta^T x)'  x_j  \\
         =  -\alpha(1-\alpha) x_k  x_j  \\
      \end{align*}
   \end{minipage}%
   \]

   Put that all back together and see if you can simplify.
   \begin{align*}
      = - \frac{1}{m} \left[0 - \alpha(1-\alpha) x_k  x_j \right]  \\
      = \frac{1}{m}\alpha(1-\alpha) x_k  x_j
   \end{align*}

   Turns out we really can't simplify it much... But that's OK. This now is in
   the form of a quadratic.  (For more information check out: TODO link
   kahn academy or explain)

   \begin{align*}
      \diffp{J(\theta)}{{\theta_j}{\theta_k}} =
      \frac{1}{m}\sum_{i=1}^{m}\alpha(1-\alpha) x_k  x_j
   \end{align*}
To change this to matrices we would get
(\href{https://zief0002.github.io/matrix-algebra/quadratic-form-of-a-matrix.html#quadratic-form}{Here is more info on this}.):
\begin{align*}
   \diffp{J(\theta)}{{\theta_j}{\theta_k}} = x_j^T
   \frac{1}{m}[\alpha(1-\alpha) ] x_k
\end{align*}
```