Name:	Student ID:

CONVEXITY PROBLEM SHEET

1

(a) Starting from the definition of a convex function

$$f(ax + (1-a)y) \leq a f(x) + (1-a) f(y) \tag{1}$$

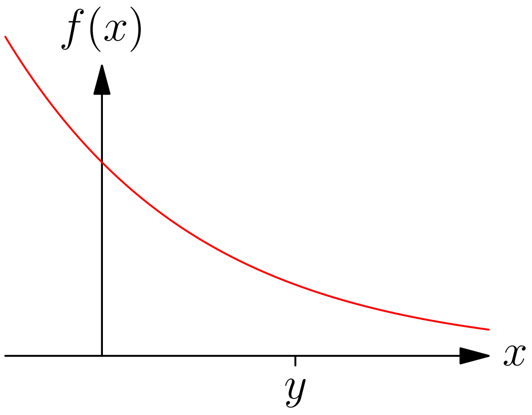
Let $a = \epsilon/(x-y)$ and rearrange the inequality to give

$$(x-y)\left(\frac{f(y+\epsilon) - f(y)}{\epsilon}\right)$$

on the left-hand side. Taking the limit $\epsilon \to 0$ show that the function $f(x)$ lies above the tangent line $t(x) = f(y) + (x-y)f'(y)$ going through the point y. [4 marks]

$\overline{4}$

(b) Sketch the tangent line, $t(x)$, at the point y in the graph shown below. [1 mark]



$\frac{}{1}$

(c) Starting from the inequality for a convex function

$$f(x) \geq f(y) + (x - y) f'(y) \tag{2}$$

consider the case $y = x + \epsilon$, then by Taylor expanding $f(x + \epsilon)$ and $f'(x + \epsilon)$ around x and keeping all terms up to order ϵ^2 show that for a convex function $f''(x) \geq 0$. [4 marks]

$\frac{}{4}$

(d) Prove that $-\log(x)$ is convex-up for $x > 0$. [1 mark]

$\frac{}{1}$

End of question 1

(a) $\frac{}{4}$ (b) $\frac{}{1}$ (c) $\frac{}{4}$ (d) $\frac{}{1}$ Total $\frac{}{10}$

2

(a) If $\|\boldsymbol{x}\|$ is a proper norm use the triangular inequality ($\|\boldsymbol{x}+\boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$), linearity of a norm ($\|a\boldsymbol{x}\| = a\|\boldsymbol{x}\|$) and the definition of convexity, to show that the norm is convex. [5 marks]

$\overline{5}$

(b) Consider a classification problem where $\hat{f}_c(\boldsymbol{x}|\boldsymbol{\theta})$ is the probability that a learning machine with parameters $\boldsymbol{\theta}$ predicts that input $\boldsymbol{x}$ belongs to class $c \in \mathcal{C}$. Assume the training is stochastic so the probability of obtaining parameters $\boldsymbol{\theta}$ is $\rho(\boldsymbol{\theta})$. Let $\hat{m}_c(\boldsymbol{x}) = \mathbb{E}_{\boldsymbol{\theta}}\left[\hat{f}_c(\boldsymbol{x}|\boldsymbol{\theta})\right]$ be the output of the mean machine for class c. Assuming that for a data point $(\boldsymbol{x},y)$, where y is a class label, we use a cross entropy loss

$$L(\boldsymbol{x},y,\boldsymbol{\theta}) = -\sum_{c \in \mathcal{C}} [\![y = c]\!] \log\left(\hat{f}_c(\boldsymbol{x}|\boldsymbol{\theta})\right),$$

show that the expected loss over inputs and parameters can be written as the expected loss of the mean machine plus a second loss. Use Jensen's inequality ($\mathbb{E}\left[\log(X)\right] \leq \log\left(\mathbb{E}\left[X\right]\right)$) to show the second term is positive. [5 marks]

$\overline{5}$

End of question 2

(a) $\overline{5}$ (b) $\overline{5}$ Total $\overline{10}$

