Problem 2

Xinyue Chen

(a) The hard-margin optimization problem is

$min_w \ \|w\|^2$, s.t. $y^{(i)}(w \cdot x^{(i)} + b) \geq 1$ for $i = 1, 2, ..., m$

Since w, b is optimal for this problem, and we derived from class that $\delta^{(i)} = \frac{y^{(i)}(w \cdot x^{(i)}+b)}{\|w\|}, \forall i,$
where $\delta^{(i)}$ is the distance from the hyperplane to the i^{th} training example.

Denote the nearest training example to hyperplane by $x^{(j)}, y^{(j)}, j \in \forall i.$
Since w, b is optimal for this problem, intuitively, w, b meets the constraints by equality on $x^{(j)}, y^{(j)}$, i.e., $y^{(j)}(w \cdot x^{(j)} + b) = 1$.
Thus $min\delta^{(i)} = \delta^{(j)} = \frac{y^{(j)}(w \cdot x^{(j)}+b)}{\|w\|} = \frac{1}{\|w\|}$, i.e., the margin for w, b (distance from hyperplane to nearest training example) is $\frac{1}{\|w\|}$.

(b) Since z, d is any other separable hyperplane and M is its margin for the data set, analogously with what we derived in class,
$\frac{y^{(i)}(z \cdot x^{(i)}+d)}{\|z\|} \geq M$ for $i = 1, 2, ..., m$ (1)

From $z' = \frac{z}{\|z\|M}$, $d' = \frac{d}{\|z\|M}$ we derive

$z = \|z\|Mz'$ (2)

$d = \|z\|Md'$ (3)
Substitude (2), (3) into (1), we get

$\frac{y^{(i)}(\|z\|Mz' \cdot x^{(i)}+\|z\|Md')}{\|z\|} \geq M$

$\Rightarrow y^{(i)}(z' \cdot x^{(i)} + d') \geq 1$

Thus z', d' is a feasible solution for the hard-margin optimization problem, and therefore $\|w\|^2 \leq \|z'\|^2$, and hence $\|w\| \leq \|z'\|$.

(c) From (a), margin for w, b is $\frac{1}{\|w\|}$; from (b), $\|w\| \leq \|z'\|$,
therefore $\frac{1}{\|w\|} \geq \frac{1}{\|z'\|}$.

From equation(2) in (b) we derive
$M = \frac{z}{\|z\|z'}$
Since z and z' points to the same direction (z' is obtained by scaling z), we can replace $\frac{z}{\|z\|}$ with $\frac{z'}{\|z'\|}$,
$M = \frac{z'}{\|z'\|z'} = \frac{1}{\|z'\|}$

Thus $\frac{1}{\|w\|} \geq M$.