Problem 1. Consider the training set $P$ of points shown below:

- $A(−1, −5)$
- $B(−3, 1)$
- $C(4, −2)$
- $D(1, 4)$

where the two dots have label 1, the cross has label 2, and the box has label 3. Run multiclass Perceptron to find a generalized linear classifier to separate $P$.

Answer: At the beginning, $\vec{w}_1 = \vec{w}_2 = \vec{w}_3 = [0, 0]$.

Round 1: Violation point $D$, $\ell = 2$, $z = 1$. Hence, $\vec{w}_1 = [−1, −4], \vec{w}_2 = [1, 4], \vec{w}_3 = [0, 0]$.
Round 2: Violation point $B$, $\ell = 3$, $z = 2$. Hence, $\vec{w}_1 = [−1, −4], \vec{w}_2 = [4, 3], \vec{w}_3 = [−3, 1]$.
Round 3: Violation point $C$, $\ell = 1$, $z = 2$. Hence, $\vec{w}_1 = [3, −6], \vec{w}_2 = [0, 5], \vec{w}_3 = [−3, 1]$.

No more violations.

Problem 2. Calculate the margin of the classifier you obtained in the previous problem.

Answer: Let $W$ be the set of weight vectors obtained.

- $\text{margin}(A \mid W) = \frac{19}{\sqrt{2\times80}}$
- $\text{margin}(B \mid W) = \frac{38}{\sqrt{2\times80}}$
- $\text{margin}(C \mid W) = \frac{29}{\sqrt{2\times80}}$
- $\text{margin}(D \mid W) = \frac{5}{\sqrt{2\times80}}$

Therefore, the margin equals $\frac{5}{\sqrt{2\times80}}$.

Problem 3. Suppose we run multiclass Perceptron on $k = 2$. Let $\{\vec{w}_1, \vec{w}_2\}$ be the set of weight vectors returned. Prove: $\vec{w}_1 = −\vec{w}_2$.

Answer: It suffices to prove that $\vec{w}_1 + \vec{w}_2 = \vec{0}$ after every round. This obviously holds at the beginning because $\vec{w}_1 = \vec{w}_2 = \vec{0}$. Suppose that $\vec{w}_1 + \vec{w}_2 = \vec{0}$ before the next round starts. Let $p$ be the violation point used in the round to do adjustments. Since we always add $\vec{p}$ to a weight vector but subtract $\vec{p}$ from the other weight vector, $\vec{w}_1 + \vec{w}_2$ is still $\vec{0}$ at the end of the round.

Problem 4. Continuing on Problem 3, prove: the “margin” of $W = \{\vec{w}_1, \vec{w}_2\}$ as defined in multiclass Perceptron is precisely the “margin” as defined in (the traditional) Perceptron (i.e., the smallest distance from a point in the training set $P$ to the separation plane).

Answer: It suffices to prove: for each point $p$ in the training set, $\text{margin}(p \mid W)$ is precisely the distance from $p$ to the separation plane.
Without loss of generality, assume that $p$ is classified as class 1, i.e., $\vec{w}_1 \cdot \vec{p} > \vec{w}_2 \cdot \vec{p}$. We have:

$$
\text{margin}(p \mid W) = \frac{\vec{w}_1 \cdot \vec{p} - \vec{w}_2 \cdot \vec{p}}{\sqrt{2(|\vec{w}_1|^2 + |\vec{w}_2|^2)}}
$$

$$
= \frac{2\vec{w}_1 \cdot \vec{p}}{\sqrt{4|\vec{w}_1|^2}}
$$

$$
= \frac{\vec{w}_1 \cdot \vec{p}}{|\vec{w}_1|}
$$

which is the distance from $p$ to the separation plane, as promised.