Perceptron Learning - Why it works?

What is the intuition behind the Perceptron learning algorithm

1. $W = [w_1, w_2, \ldots w_n]$
2. $X = [x_1, x_2, \ldots x_n]$
3. $\cos\theta = w.x/||w||x||$, here the numerator can be replaced with $\Sigma\, w_i x_i$
   a. The denominator is always positive
   b. Therefore $\cos\theta \propto \Sigma\, w_i x_i$
   c. As θ ranges from 0 to 180°, $\cos\theta$ ranges from 1 to -1
   d. If $\cos\theta > 0$, it is an acute angle
   e. If $\cos\theta < 0$, it is an obtuse angle
4. For $x \in P$, if $w.x < 0$, then it means that the angle(α) between this x and the current w is greater than 90°, but we want $\alpha < 90°$)
   a. What happens to the new angle α_{new} when $w_{new} = w + x$
   b. $\cos\alpha_{new} \propto {w_{new}}^T x$
   c. $\propto (w+x)^T x$
   d. $\propto w^T x + x^T x$(always +ve)
   e. $\propto \cos\alpha + x^T x$(some +ve value)
   f. This means that the cosine is going to increase, which leads to decrease of α
5. For $x \in N$, if $w.x > 0$, then it means that the angle(α) between this x and the current w is less than 90°, but we want $\alpha > 90°$)
   a. What happens to the new angle α_{new} when $w_{new} = w - x$
   b. $\cos\alpha_{new} \propto {w_{new}}^T x$
   c. $\propto (w-x)^T x$
   d. $\propto w^T x - x^T x$(always +ve)
   e. $\propto \cos\alpha - x^T x$(some +ve value)
   f. This means that the cosine is going to decrease, which leads to increase of α