Regularization is designed to smoothen the problem objective and also prevents overfitting.

Why does it help with overfitting?

Consider two data points x and x+ $\Delta x$ and suppose both have same labels.

Now consider their prediction: $w^T x + w_0$ and $w^T(x + \Delta x) + w_0$. Since the two points are close we want their predictions to also be similar or close. Imagine if w had very large numbers. Then the second term would become $w^T x + w^T \Delta x + w_0 = w^T x + w_0 + w^T \Delta x$ . We want this last term $w^T \Delta x$ to be small and that can be done if the entries of w are small. If the entries of w are very large then there is no guarantee.