Although there is a great success of applying deep learning on a wide variety of tasks, it heavily relies on a large amount of labeled training data, which could be hard to obtain in many real scenarios. To address this problem, unsupervised and semi-supervised learning emerge to take advantage of the plenty of cheap unlabeled data to improve the model generalization. In this dissertation, we present methods on applying various forms of perturbations to approach robust unsupervised and semi-supervised learning. Features are learnt via auto-encoding the perturbations on the input data. Models are regularized through minimizing the effects of perturbations on features and model parameters. The idea is as follows: the features of a robust model ought to be sufficiently informative and equivariant to perturbations on the input data, and the classifiers should be resilient and invariant to small perturbations on features and model parameters. Experiments on several benchmarks show the proposed methods outperform many state-of-the-art approaches on unsupervised and semi-supervised learning, proving the perturbation equivariance and invariance are critical principles for robust feature representation learning.