Three important properties of a classification machinery are i) the system preserves the core information of the input data; ii) the training examples convey information about unseen data; and iii) the system is able to treat differently points from different classes. In this paper, we show that these fundamental properties are satisfied by the architecture of deep neural networks. We formally prove that these networks with random Gaussian weights perform a distance-preserving embedding of the data, with a special treatment for in-class and out-of-class data. Similar points at the input of the network are likely to have a similar output. The theoretical analysis of deep networks here presented exploits tools used in the compressed sensing and dictionary learning literature, thereby making a formal connection between these important topics. The derived results allow drawing conclusions on the metric learning properties of the network and their relation to its structure, as well as providing bounds on the required size of the training set such that the training examples would represent faithfully the unseen data. The results are validated with state-of-the-art trained networks.