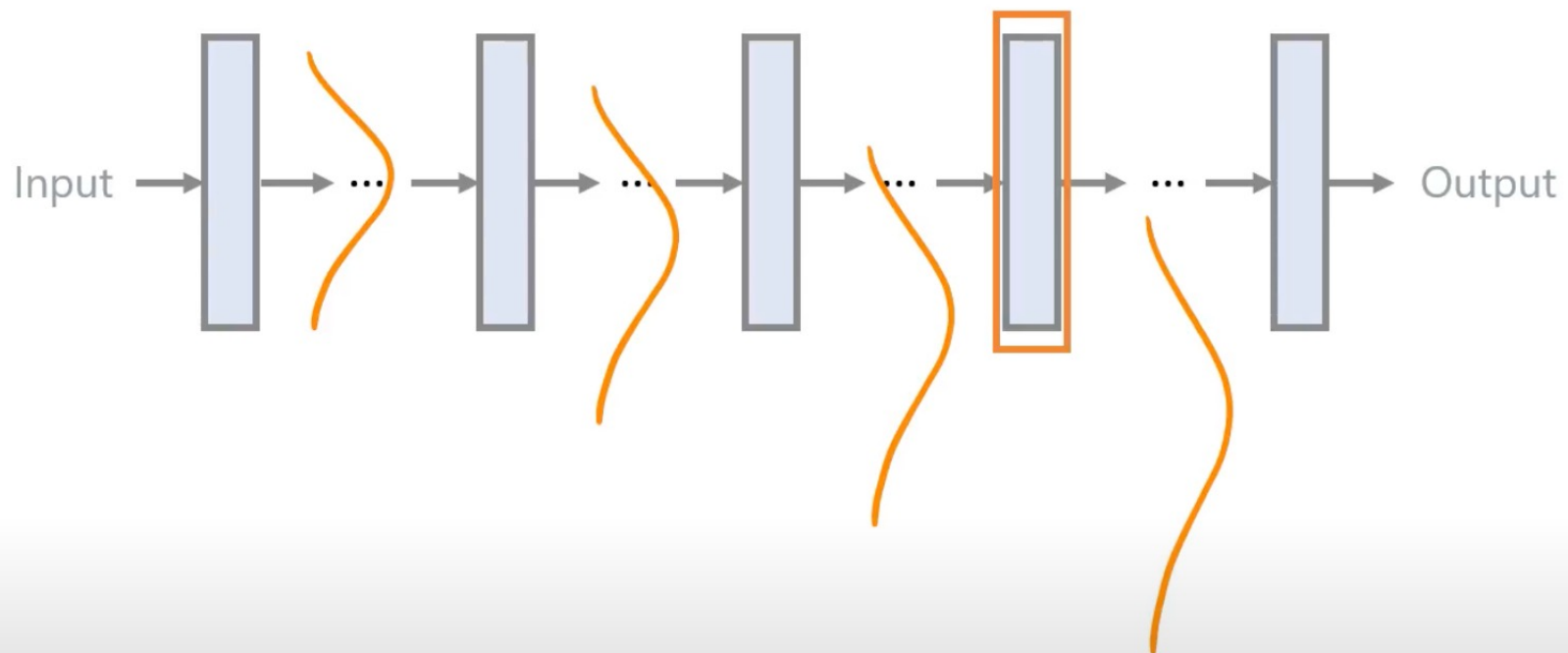
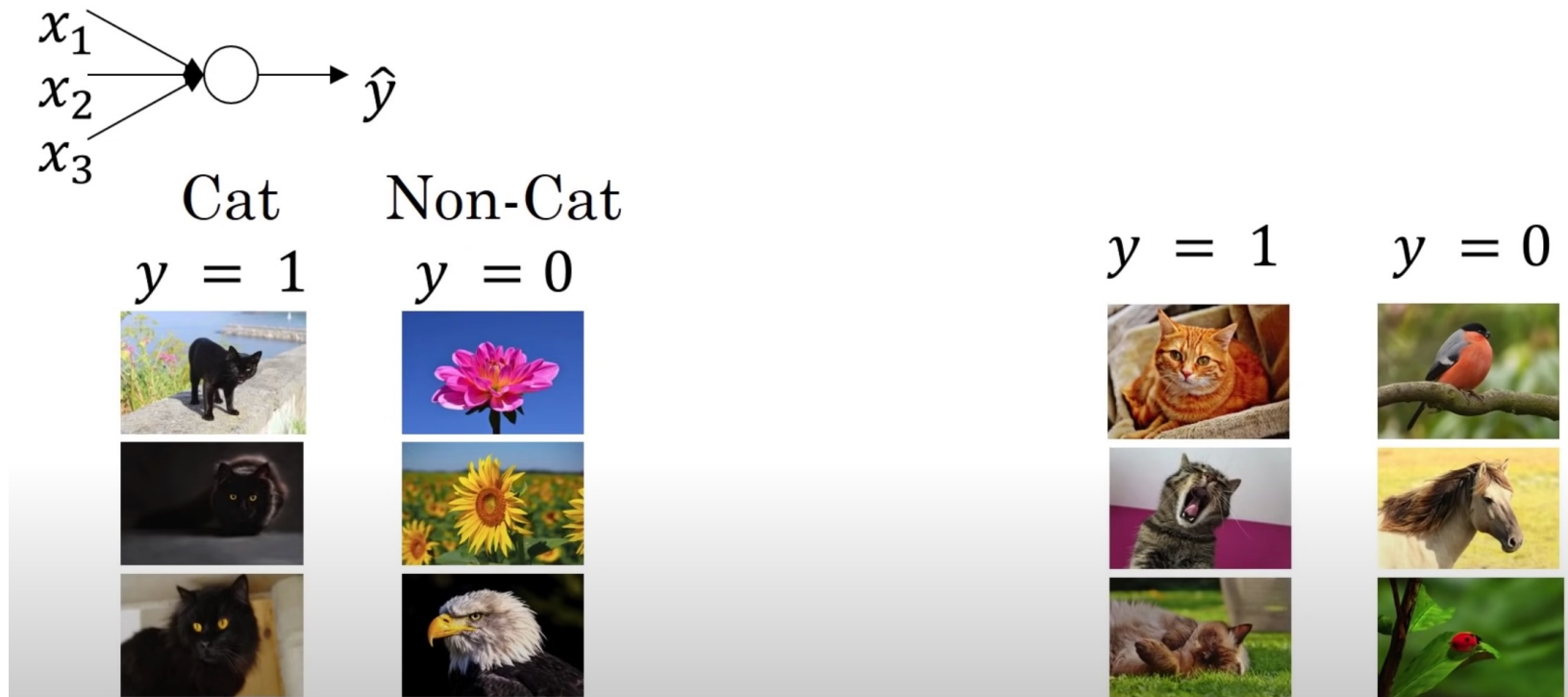


The situation without batch normalization, network activations are exposed to varying data input distributions that propagate through the network and distort the learned distributions.

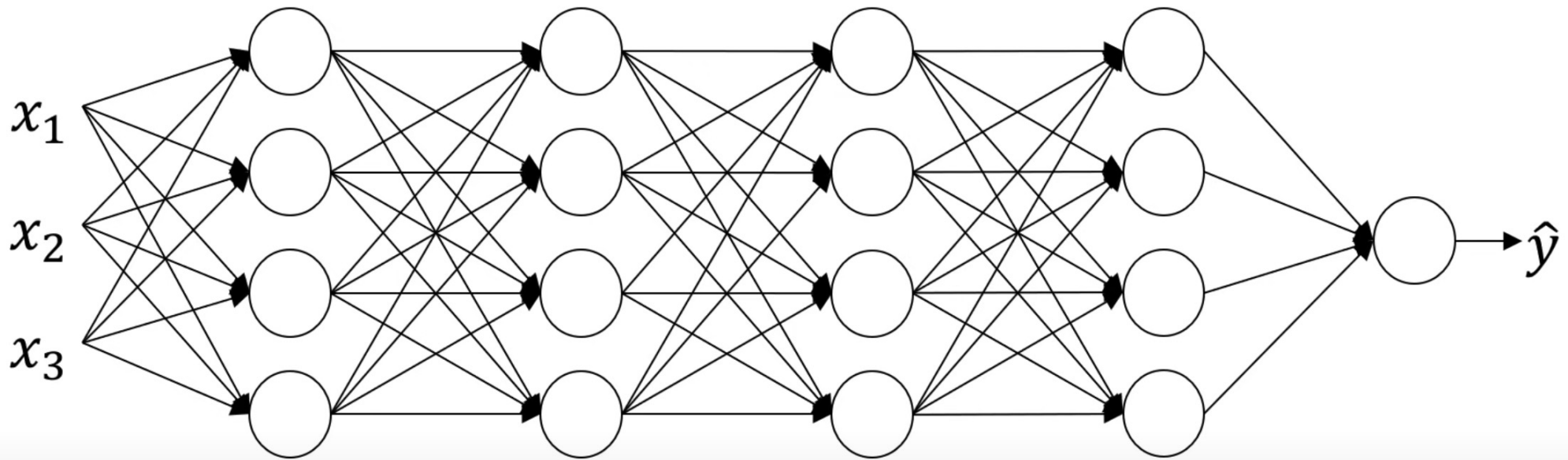
<https://towardsdatascience.com/understanding-dataset-shift-f2a5a262a766>

Internal covariate shift





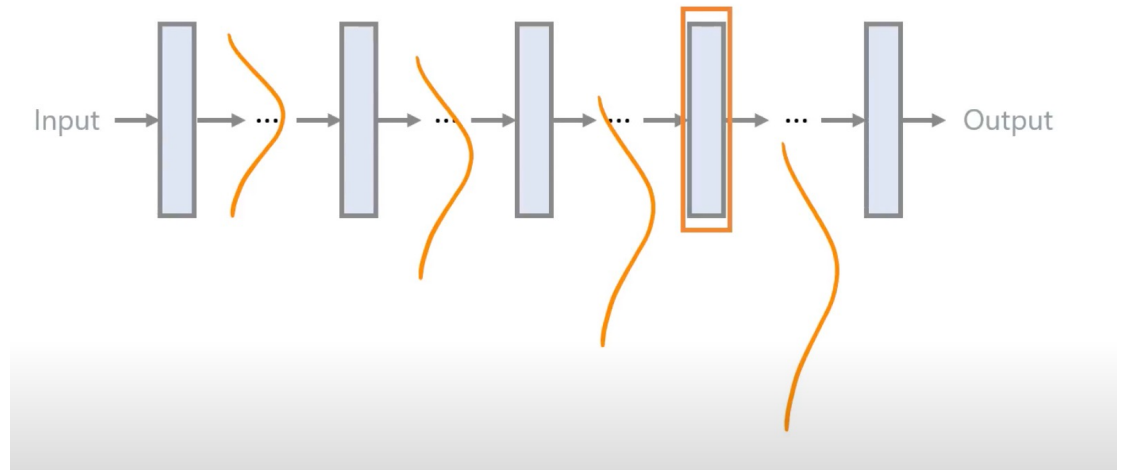
En utilisant la normalisation par lots, le décalage des valeurs unitaires cachées (c'est-à-dire le décalage de covariable – internal covariate shift) est réduit dans une large mesure. Un exemple de changement de « covariate » peut-être celui-ci : disons que nous avons préparé et entraîné un modèle qui nous indique s'il y a un chat dans l'image d'entrée. S'il y a un chat, alors la sortie est 1 sinon 0. Maintenant, si l'ensemble d'entraînement ne contient que les images de chats noirs et pour la prédiction, nous avons donné une entrée comme l'image d'un chat de couleur, alors notre modèle ne fonctionnerait pas avec une bonne efficacité. C'est ce que signifie fondamentalement le changement de covariables.



Un décalage des covariables interne (internal covariate shift) se produit **lorsqu'il y a un changement dans la distribution d'entrée de notre réseau. Lorsque la distribution d'entrée change, les couches cachées essaient d'apprendre à s'adapter à la nouvelle distribution.** Cela ralentit le processus d'apprentissage. Si un processus ralentit, il faut beaucoup de temps pour converger vers un minimum global. Ce problème se produit lorsque la distribution statistique de l'entrée aux réseaux est bien différente de l'entrée qu'elle a vue auparavant. **La normalisation par lots et d'autres techniques de normalisation peuvent résoudre ce problème.**

Internal Covariate Shift can be defined as the change in the distribution of network activations due to the change in network parameters during training.

Internal covariate shift



Normalization

