<table>
<tr><th></th><th>doc_1</th><th>doc_2</th><th>decision</th><th>id</th></tr>
<tr><th>cases</th>
<td><table>
<tr><th>authors</th><td>Gorur, Dilan
Lakshminarayanan, Balaji
Matsukawa, Akihiro
Nalisnick, Eric
Teh, Yee Whye</td></tr>
<tr><th>title</th><td>Do Deep Generative Models Know What They Don't Know?</td></tr>
<tr><th>publication_date</th><td>2019-01-01 00:00:00</td></tr>
<tr><th>source</th><td>SupportedSources.CORE</td></tr>
<tr><th>journal</th><td></td></tr>
<tr><th>volume</th><td></td></tr>
<tr><th>doi</th><td>None</td></tr>
<tr><th>urls</th><td>http://arxiv.org/abs/1810.09136</td></tr>
<tr><th>id</th><td>id3554507340361278977</td></tr>
<tr><th>abstract</th><td>A neural network deployed in the wild may be asked to make predictions for inputs that were drawn from a different distribution than that of the training data. A plethora of work has demonstrated that it is easy to find or synthesize inputs for which a neural network is highly confident yet wrong. Generative models are widely viewed to be robust to such mistaken confidence as modeling the density of the input features can be used to detect novel, out-of-distribution inputs. In this paper we challenge this assumption. We find that the density learned by flow-based models, VAEs, and PixelCNNs cannot distinguish images of common objects such as dogs, trucks, and horses (i.e. CIFAR-10) from those of house numbers (i.e. SVHN), assigning a higher likelihood to the latter when the model is trained on the former. Moreover, we find evidence of this phenomenon when pairing several popular image data sets: FashionMNIST vs MNIST, CelebA vs SVHN, ImageNet vs CIFAR-10 / CIFAR-100 / SVHN. To investigate this curious behavior, we focus analysis on flow-based generative models in particular since they are trained and evaluated via the exact marginal likelihood. We find such behavior persists even when we restrict the flows to constant-volume transformations. These transformations admit some theoretical analysis, and we show that the difference in likelihoods can be explained by the location and variances of the data and the model curvature. Our results caution against using the density estimates from deep generative models to identify inputs similar to the training distribution until their behavior for out-of-distribution inputs is better understood.Comment: ICLR 201</td></tr>
<tr><th>versions</th><td></td></tr>
</table></td>
<td><table>
<tr><th>authors</th><td>Eric Nalisnick
Balaji Lakshminarayanan
Dilan Gorur
Yee Whye Teh
Akihiro Matsukawa</td></tr>
<tr><th>title</th><td>Do Deep Generative Models Know What They Don't Know?</td></tr>
<tr><th>publication_date</th><td>2018-10-22 00:00:00</td></tr>
<tr><th>source</th><td>SupportedSources.PAPERS_WITH_CODE</td></tr>
<tr><th>journal</th><td></td></tr>
<tr><th>volume</th><td></td></tr>
<tr><th>doi</th><td></td></tr>
<tr><th>urls</th><td>http://arxiv.org/pdf/1810.09136v3.pdf
https://github.com/glouppe/info8010-deep-learning
https://openreview.net/pdf?id=H1xwNhCcYm</td></tr>
<tr><th>id</th><td>id-1249127631546035503</td></tr>
<tr><th>abstract</th><td>A neural network deployed in the wild may be asked to make predictions for inputs that were drawn from a different distribution than that of the training data. A plethora of work has demonstrated that it is easy to find or synthesize inputs for which a neural network is highly confident yet wrong. Generative models are widely viewed to be robust to such mistaken confidence as modeling the density of the input features can be used to detect novel, out-of-distribution inputs. In this paper we challenge this assumption. We find that the density learned by flow-based models, VAEs, and PixelCNNs cannot distinguish images of common objects such as dogs, trucks, and horses (i.e. CIFAR-10) from those of house numbers (i.e. SVHN), assigning a higher likelihood to the latter when the model is trained on the former. Moreover, we find evidence of this phenomenon when pairing several popular image data sets: FashionMNIST vs MNIST, CelebA vs SVHN, ImageNet vs CIFAR-10 / CIFAR-100 / SVHN. To investigate this curious behavior, we focus analysis on flow-based generative models in particular since they are trained and evaluated via the exact marginal likelihood. We find such behavior persists even when we restrict the flows to constant-volume transformations. These transformations admit some theoretical analysis, and we show that the difference in likelihoods can be explained by the location and variances of the data and the model curvature. Our results caution against using the density estimates from deep generative models to identify inputs similar to the training distribution until their behavior for out-of-distribution inputs is better understood.</td></tr>
<tr><th>versions</th><td></td></tr>
</table></td>
<td>DUPLICATES</td>
<td>164</td></tr>
</table>