<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>Wesley J. Maddox
Timur Garipov
Pavel Izmailov
Dmitry Vetrov
Andrew Wilson</td></tr>
<tr><th>title</th><td>A Simple Baseline for Bayesian Uncertainty in Deep Learning</td></tr>
<tr><th>publication_date</th><td>2019-02-07 00:00:00</td></tr>
<tr><th>source</th><td>SupportedSources.OPENALEX</td></tr>
<tr><th>journal</th><td>arXiv (Cornell University)</td></tr>
<tr><th>volume</th><td></td></tr>
<tr><th>doi</th><td>None</td></tr>
<tr><th>urls</th><td>https://openalex.org/W2912168444</td></tr>
<tr><th>id</th><td>id8577650753981608608</td></tr>
<tr><th>abstract</th><td></td></tr>
<tr><th>versions</th><td></td></tr>
</table></td>
<td><table>
<tr><th>authors</th><td>Pavel Izmailov
Timur Garipov
Andrew Gordon Wilson
Dmitry Vetrov
Wesley Maddox</td></tr>
<tr><th>title</th><td>A Simple Baseline for Bayesian Uncertainty in Deep Learning</td></tr>
<tr><th>publication_date</th><td>2019-02-07 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>https://arxiv.org/pdf/1902.02476v2.pdf
https://github.com/wjmaddox/swa_gaussian
http://papers.nips.cc/paper/9472-a-simple-baseline-for-bayesian-uncertainty-in-deep-learning.pdf</td></tr>
<tr><th>id</th><td>id2036108193044549007</td></tr>
<tr><th>abstract</th><td>We propose SWA-Gaussian (SWAG), a simple, scalable, and general purpose approach for uncertainty representation and calibration in deep learning. Stochastic Weight Averaging (SWA), which computes the first moment of stochastic gradient descent (SGD) iterates with a modified learning rate schedule, has recently been shown to improve generalization in deep learning. With SWAG, we fit a Gaussian using the SWA solution as the first moment and a low rank plus diagonal covariance also derived from the SGD iterates, forming an approximate posterior distribution over neural network weights; we then sample from this Gaussian distribution to perform Bayesian model averaging. We empirically find that SWAG approximates the shape of the true posterior, in accordance with results describing the stationary distribution of SGD iterates. Moreover, we demonstrate that SWAG performs well on a wide variety of tasks, including out of sample detection, calibration, and transfer learning, in comparison to many popular alternatives including MC dropout, KFAC Laplace, SGLD, and temperature scaling.</td></tr>
<tr><th>versions</th><td></td></tr>
</table></td>
<td>DUPLICATES</td><td>156</td></tr>
</table>