**Jawaban**

1. Ada pada file program : [simple\_PSO.ipynb]
2. Ada pada file program : [simple\_NN.ipynb], Penjelasan :
3. Final values for all node on hidden layer after 1-time forward propagation:

Jawab : Final values utk semua node = [0.99997246], [0.97068777], [0.99999917]

1. Mean square error (MSE) after forward propagation :

Jawab : MSE terakhir = [0.00028241393298124843]

1. Updated weights after 1-time backpropagation : [7.0001464]
2. Mean square error (MSE) after backpropagation : [2.1744296832333454]
3. After conducting simulation based on the description above, what can you conclude? Is the neural network successful in conducting learning process?  
   Jawab : Kesimpulan pada proses NN ini ialah Neural Networks didasarkan pada kumpulan unit yang terhubung (neuron), yang, seperti sinapsis di otak, dapat mengirimkan sinyal ke neuron lain, sehingga, bertindak seperti sel otak yang saling berhubungan, mereka dapat belajar dan membuat keputusan dalam waktu yang lebih lama. cara seperti manusia. Maka neural network bisa dikatakan cukup berhasil (belum sempurna) melakukan proses pembelajaran.
4. Apakah memungkinkan bahwa batch normalization dapat membantu proses optimasi? Silakan dapat memberikan pandangan dan tanggapan berbasis pada penelitian berikut: (S. Santurkar, D. Tsipras, A. Ilyas, A. Madry. How does batch normalization help optimization? Advances in neural information processing systems, 2018).

Jawab : Normalisasi batch (BatchNorm) merupakan arsitektur yang efektif dalam deep learning. Alasannya pada saat proses training membuat parameter ulang

mendasari masalah optimasi untuk membuatnya lebih stabil (dalam arti loss Lipschitzness) dan smooth (dalam arti "efektif" β-smoothness of the loss). Artinya gradien yang digunakan dalam pelatihan lebih prediktif dan berperilaku baik, yang memungkinkan pengoptimalan yang lebih cepat dan lebih efektif.

1. Berikut ini merupakan arsitektur dense dengan jumlah layer sebanyak 5, serta growth rate k = 4. Setiap layer pada arsitektur ini mengambil seluruh feature map sebelumnya sebagai input. (G. Huang, Z. Liu, K. Q. Weinberger and L. Maaten. Densely Connected Convolutional Networks. arXiv:1608.06993v3, 2016). Kemukakan ulasan Anda terhadap arsitektur tersebut dibandingkan dengan vanilla Convolutional Neural Networks.

Jawab : jaringan konvolusional bisa menjadi jauh lebih dalam, lebih akurat, dan efisien untuk training jika mengandung koneksi yang lebih pendek antar lapisan yang dekat dengan input dan output. Dense Convolutional Network (DenseNet), yang menghubungkan setiap lapisan ke setiap lapisan lainnya dengan cara feed- forward fashion. sedangkan Sedangkan jaringan konvolusional tradisional dengan lapisan L memiliki L koneksi — satu di antara setiap lapisan dan lapisan berikutnya memilki koneksi langsung. feature-maps dari semua lapisan sebelumnya adalah

digunakan sebagai input, dan feature-maps sendiri digunakan sebagai input

ke semua lapisan berikutnya. DenseNets memiliki beberapa keuntungan menarik: meringankan masalah vanishing-gradient memperkuat feature propagation, mendorong penggunaan kembali feature dan secara substansial mengurangi jumlah parameter. Dense terdiri dari 5 layer, growth rate k=4 dan setiap layer pada arsitektur ini mengambil seluruh feature map sebelumnya sebagai input.

Vanilla CNN terdiri dari 3 convolution layers. Di setiap lapisan, filter berukuran lima

digunakan. Jumlah filter pertama, kedua, dan ketiga lapisan adalah 8, 16, dan 120.

pada penelitian [1], Dense-Net dibandingkan dengan vanilla CNN atau semua jenis CNN memiliki hasil eror yang paling kecil hingga 1.59% . Semua hasil DenseNets tanpa augmentasi data (C10, C100, SVHN) diperoleh dengan menggunakan Dropout.

1. Soal-soal berikut terkait dengan penerapan Deep Learning pada bidang Text Analysis:
2. [LO1, LO2, LO4, LO5, Bobot 18%] Jelaskan dengan lengkap arsitektur dan cara kerja dari model BERT yang diusulkan oleh Devlin dkk (<https://arxiv.org/abs/1810.04805>). Penjelasan yang diberikan harus dalam Bahasa Indonesia.

Jawab : Arsitektur BERT ialah multi-layer bidirectional Transformer encoder, dimana terdiri dari 2 kerangka kerja, yaitu pre-training and fine-tuning. selama pre-training, model dilatih pada data yang tidak berlabel melalui berbagai tugas pre-training. Untuk fine- tuning, model BERT pertama kali diinisialisasi dengan parameter pre-trained. dan semua parameter di fine-tuned menggunakan data berlabel dari downstream tasks. setiap downstream task memiliki model fine-tuned terpisah, meskipun mereka diinisialisasi dengan parameter pre-trained yang sama.

Cara kerjanya : 1. Pre-training BERT : kami melakukan pre-train BERT menggunakan dua unsupervised tasks.

* Masked LM : Untuk melatih deep bidirectional representation, dengan menutupi beberapa persentase dari token input secara acak, dan kemudian memprediksi token yang di masked. vektor tersembunyi terakhir yang sesuai dengan mask tokens dimasukkan ke softmax keluaran melalui kosakata, seperti dalam standar LM.
* Next Sentence Prediction (NSP) : melakukan pre-train untuk tugas prediksi kalimat binerisasi berikutnya yang dapat dibuat dengan mudah dari korpus monolingual mana pun. Pre-training data : Untuk korpus Pre-training menggunakan bookscorpus dan wikipedia bahasa inggris untuk mengekstrak urutan panjang yang berdekatan.

2. Fine-tuning BERT : mekanisme di Transformer memungkinkan BERT untuk memodelkan banyak downstream tasks - dengan menukar input dan output yang sesuai. Untuk setiap tugas, kami cukup menyambungkan input dan output khusus tugas ke dalam BERT dan menyempurnakan semua parameter end-to-end. Pada bagian input, kalimat A dan kalimat B hasil pre-training dianalogikan dengan (1) pasangan kalimat dalam parafrase, (2) pasangan hipotesis-premis dalam ensertamen, (3) pasangan pertanyaan-bagian dalam tanya jawab, dan

(4) pasangan teks-∅ yang merosot dalam klasifikasi teks atau penandaan urutan. Pada keluaran, representasi token dimasukkan ke lapisan keluaran untuk tugas tingkat token, seperti penandaan urutan atau penjawaban pertanyaan, dan representasi [CLS] dimasukkan ke dalam lapisan keluaran untuk klasifikasi, seperti entailment atau analisis sentimen.

1. [LO1, LO2, LO4, LO5, Bobot 10%] Jelaskan perbaikan-perbaikan dari BERT yang diusulkan pada RoBERTa oleh Liu dkk (https://arxiv.org/abs/1907.11692). Penjelasan yang diberikan harus dalam Bahasa Indonesia.

Jawab : Perbaikan :

* melatih model lebih lama, dengan batch yang lebih besar, lebih banyak data : Untuk menghindari penggunaan mask yang sama untuk setiap instance training di setiap epoch, data training digandakan sebanyak 10 kali sehingga masing-masing urutannya adalah ditutupi dengan 10 cara berbeda selama 40 epoch training. Dengan demikian, setiap urutan pelatihan terlihat dengan mask yang sama selama pelatihan. Dengan masking statis melakukan mirip denganmodel BERT asli, dan masking dinamis sebanding atau sedikit lebih baik daripada masking statis.
* menghapus kalimat berikutnya tujuan prediksi : melakukan segment-pair+nsp, Setiap input memiliki sepasang segmen, yang masing-masing dapat berisi beberapa natural kalimat, tetapi total panjang gabungan harus kurang dari 512 token. Lalu sentence-pair+nsp, full-sentences, dan doc-sentences. Menggunakan individu kalimat merusak kinerja di downstream tasks.
* training dengan urutan yang lebih panjang : pelatihan dengan batch besar meningkatkan perplexity untuk masked language modeling dan akurasi akhir. Batch besar juga lebih mudahmemparalelkan melalui pelatihan paralel data terdistribusi.
* Secara dinamis mengubah pola masking yang diterapkan pada data training: BPE mencapai sedikit kinerja tugas akhir yang lebih buruk pada beberapa tugas. keuntungan dari skema pengkodean universal melebihi penurunan kecil dalam kinerja dan menggunakan pengkodean ini disisa percobaan kami. Perbandingan yang lebih rinci dari pengkodean ini diserahkan pada pekerjaan di masa mendatang.

1. [LO3, LO4, Bobot 15%] Buat implementasi model BERT atau RoBERTa (bisa dipilih salah satu saja) untuk klasifikasi text menggunakan dataset IMDB 50K Movie Reviews. Dataset IMDB dapat diakses menggunakan torchtext.

Jawab : Ada pada file program : [SentimentAnalysis\_BERT.ipynb]

1. [LO5, LO6, 7%] Identifikasi kekurangan dari model BERT dan RoBERTa. Kemudian berikan usulan perbaikan untuk menghasilkan model klasifikasi text yang lebih baiklagi.

Jawb : Kekurangan BERT dan RoBERTa ialah besarnya data dan step proses yang bisa mencapai ratusan ribu langkah yang dimana membutuhkan sumber daya komputer untuk melakukannya. Besarnya data set yang dihasilkan pada proses training hingga lebih dari 100GB.

Referensi :

* [1] Videla, Lakshmi Sarvani, and PM Ashok Kumar. "Facial expression classification using vanilla convolution neural network." *2020 7th international conference on smart structures and systems (ICSSS)*. IEEE, 2020.