**Introduction**

**Work Flow

Description automatically generated**

**Data Description**

**A picture containing text

Description automatically generated**

**Principal components analysis**

**Graphical user interface, chart

Description automatically generated**

**Text

Description automatically generated with medium confidence**

**SVM**

**For the dataset T and Hyperplane(w,b), we define the geometric interval between sample points and Hyperplane is:**

**，**

**Then we define the smallest geometric interval as**

***,***

**Our goal is to maximize the the smallest geometric interval, the optimization problem generated：**

**Because of the relationship between geometric intervals and function intervals：**

**Then we set function interval to be 1,then we can change the problem to :**

**This is a convex quadratic function,after we apply regularization and kernel function to other problem, the optimization function changes to：**

**The scikit-learn uses coordinate descent as optimizer, after the calculation we will get**

**and our decision function：**

**In this dataset, the accuracy of this method is 88.32%**

**Logistic regression**

**(Convolutional neutral network) Resnet-50 math formula and references**

Text, letter

Description automatically generatedDiagram

Description automatically generatedhttps://medium.com/@bakiiii/microsoft-presents-deep-residual-networks-d0ebd3fe5887Text

Description automatically generated with medium confidencehttps://math.stackexchange.com/questions/78575/derivative-of-sigmoid-function-sigma-x-frac11e-x

derivative of sigmoid function is less than 1/4

w1,w2,w3,w4 are initialized using gaussian method to have a mean of 0 and standard deviation of 1.

**when inputs become very small or very large, the** **sigmoid function saturates at 0 and 1 and the tanh function saturates at -1 and 1. In both these cases, their derivatives are extremely close to 0.**

Chart, line chart

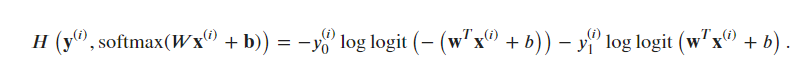
Description automatically generated

Chart, line chart

Description automatically generated

ReLU has gradient 1 when input > 0, and zero otherwise. Thus, **multiplying a bunch of ReLU derivatives together in the backprop equations has the nice property of being either 1 or 0.**

<https://medium.com/analytics-vidhya/how-batch-normalization-and-relu-solve-vanishing-gradients-3f1a8ace1c88>



Categorical cross entropy is used as the loss function for Resnet algorithm. Although it is a composition of many convex functions. However, it is a non-convex function. Since there are no-linear function as activation function for multiple layers.

We can find local minimal for the non-convex function. Therefore, the loss function is convergence for local minimum..

Text

Description automatically generated

For ResNet-50 model, it is a convolutional neural network

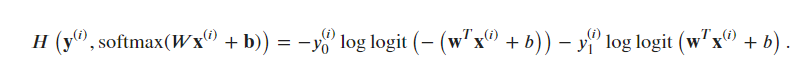
Diagram

Description automatically generatedText

Description automatically generated with medium confidence

1. Solve degradation problem in propagate processing
2. Solve vanishing gradient by using identity mappings

The loss function is categorical cross entropy



Non-Convex but convergent for local minimum.

**Activation: Rectified linear unit (ReLu)** function: max(0, x)

**Optimization approach:** Adam (a first-order gradient-based algorithm for stochastic object function) used as optimization

The accuracy of ResNet-50 [convolutional neural network model is over](https://arxiv.org/abs/1512.03385) **[90%](https://arxiv.org/abs/1512.03385)**

Chart

Description automatically generated

References:

1. Tsuboi: Fast Newton-CG Method for Batch Learning of Conditional Random Fields Tokyo (Japan): AAAI; c2011 [updated 2011 Jan 31; cited 2021 Dec 1]. Available from: [https://www.aaai.org/ocs/index.php/AAAI/AAAI11/paper/viewFile/3448/3876#:~:text=Newton%2DCG%20methods%20are%20a,of%20the%20full%20Hessian%20matrices](https://www.aaai.org/ocs/index.php/AAAI/AAAI11/paper/viewFile/3448/3876).
2. D. P. Kingma, J. Lei Ba, ADAM: A METHOD FOR STOCHASTIC OPTIMIZATION, the 3rd International Conference for Learning Representations, San Diego(CA), 2015.Available from: <https://arxiv.org/pdf/1412.6980.pdf>
3. Fashion-MNIST Dataset, Zalando Research, 2018 Available from: <https://www.kaggle.com/zalando-research/fashionmnist>
4. Chang C-C, Lin C-J. LIBSVM: A Library for Support Vector Machines. 2019 Nov 20. Available from: [https://www.csie.ntu.edu.tw](https://www.csie.ntu.edu.tw/)
5. Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. Deep Residual Learning for Image Recognition. 2015，arXiv:1512.03385 [cs.CV] Available from: <https://arxiv.org/abs/1512.03385>