ENPM703- Assignment-1

Part3: Softmax

Softmax Classifier

Softmax is used as a multiclass classifier where the scores of each class are transformed into probabilities which are used to predict the class.

Softmax Loss

The Softmax loss function takes the output scores of individual classes and transform them to probability values. The scores are the dot product of the image features and the weight that we use for each class. Here the scores are logits and the loss function takes the log of the probabilities of the true class. Since we are dealing with the log of the probabilities, the minimum and maximum value of the output is zero to infinity.

$$\sigma = -\log(e^{sj}/ \Sigma e^{si})$$

Since the model is predicting the probability of Y given X, at initialization the model is not given with the input features the probability becomes the baseline probability. In this case, that is 0.1 thus, the loss at initialization is -log (0.1).

Gradient Calculation

Rewriting the loss function as,
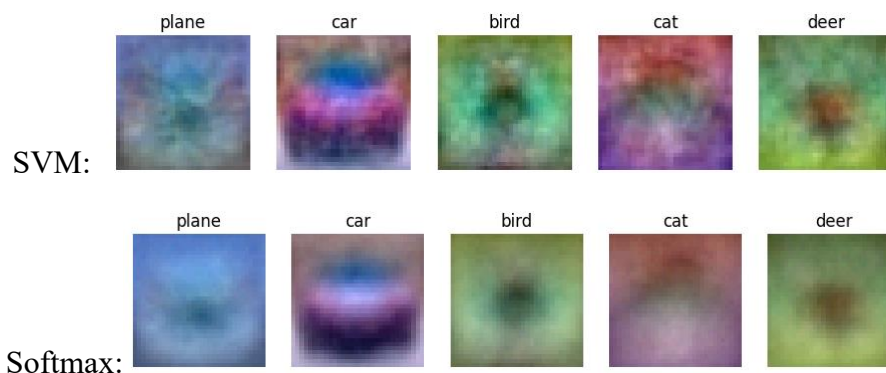
$$\sigma = - S_{correct} + \log(\Sigma e^{si})$$

Differentiating individually,

$$D(- S_{correct}) = -x_i$$

$$D(\log(\Sigma e^{si})) = \Sigma e^{si}. \, x_i$$

Finally, using cross validation the optimal values of learning rate and regularization strength are found.

SVM vs Softmax

- The visualization of weights is similar in both cases, since both weights resembles the base features of the class images. But, Softmax weights are less noisy because of its probabilistic nature and leading to smooth transition between classes.

- Since SVM uses Hinge loss to make prediction, the output is binary when compared with the probabilistic output of Softmax. Thus, Softmax can be used for multiclass classification which is often the case in real world problems.
- In SVM, the loss can approach to zero and it will not affect the solution after that point but, Softmax on the other hand will make minimal updates and will try to make the probability close to one. Thus, Softmax is more sensitive to changes.