

2 Training SVM with SGD

2.1 SGD with Momentum

The historical weight w_t for 200 time-steps using momentum parameter $\beta = 0$ and $\beta = 0.9$ was plotted as following:

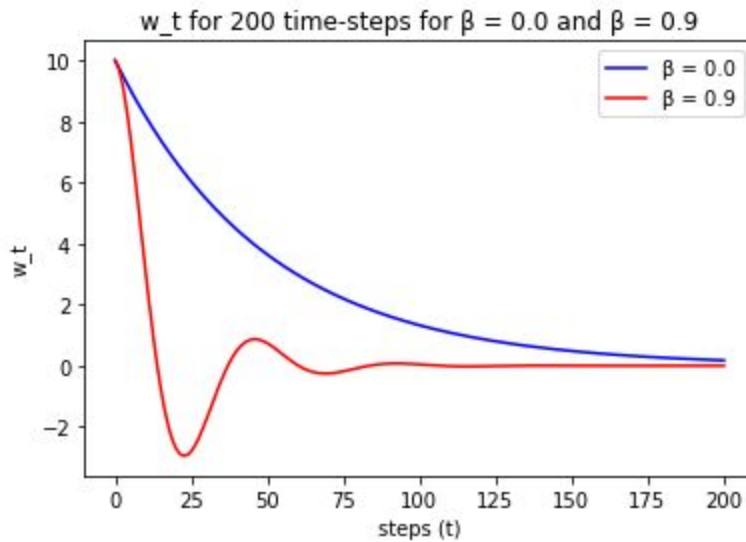


Figure 4. Plot of w_t for 200 time-steps using $\beta = 0$ and $\beta = 0.9$

As seen in the graph above, w_t with larger momentum parameter ($\beta = 0.9$) converge to a stable value of ~ 0 faster (at around 90th step) than ones with no momentum ($\beta = 0.0$; at around 200th step).

2.3 Apply on 4-vs-9 digits on MNIST

The following table summarizes the classification results in training and test set for SVM models with $\beta = 0$ and $\beta = 0.1$:

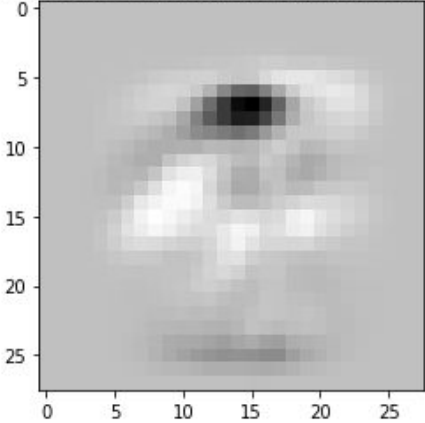
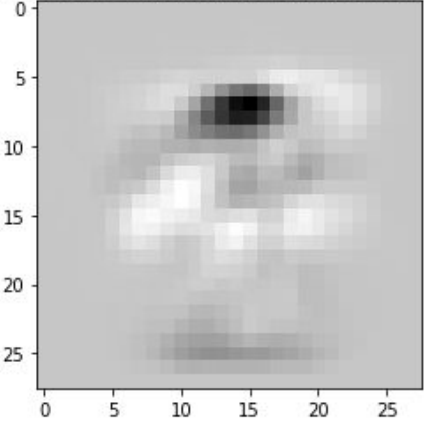
	SGD with $\beta = 0$	SGD with $\beta = 0.1$
Training Loss (Average Hinge Loss on Training Set)	0.3972404859	0.3545821809
Test Loss (Average Hinge Loss on Test Set)	0.4006276311	0.3427825853
Classification Accuracy on Training Set	0.9126530612	0.9057596371
Classification Accuracy on Test Set	0.9147624229	0.9038810301
Plot of Weights \mathbf{w} as a 28 x 28 Image	Plot of 4 vs. 9 SVM weights for momentum = 0.0 	Plot of 4 vs. 9 SVM weights for momentum = 0.1 

Table 2. Classification Results for SVM Models with $\beta = 0$ and $\beta = 0.1$