Classification loss functions:

1. Cross Entropy Loss/Negative Log Likelihood and Log loss seem to be same.
2. Hinge Loss/Multi class SVM Loss.

Logarithmic loss minimization leads to well-behaved probabilistic outputs.

Hinge loss leads to some (not guaranteed) sparsity on the dual, but it doesn't help at probability estimation. Instead, it punishes misclassifications (that's why it's so useful to determine margins): diminishing hinge-loss comes with diminishing across margin misclassifications.

So, summarizing:

* Logarithmic loss leads to better probability estimation at the cost of accuracy
* Hinge loss leads to better accuracy and some sparsity at the cost of much less sensitivity regarding probabilities

<https://towardsdatascience.com/common-loss-functions-in-machine-learning-46af0ffc4d23>

<https://medium.com/mlreview/gradient-boosting-from-scratch-1e317ae4587d>

ROC,AUC

<https://acutecaretesting.org/en/articles/roc-curves-what-are-they-and-how-are-they-used>

L1 and L2:

<https://www.analyticsvidhya.com/blog/2016/01/complete-tutorial-ridge-lasso-regression-python/>

=============================================================================

<https://www.linkedin.com/posts/srivatsan-srinivasan-b8131b_datascience-end2endds-activity-6567018330446794752-v1BX>