CS4780 Lec15 - SVM contd.

Maximize margin.

SVM
$$\begin{cases} \min_{w,b} \; w^T w \\ \text{s.t. } \forall_i \; y_i(w^T x_i + b) \geq 1 \end{cases}$$

When such a hyperplane is not feasible, add a "slack" here

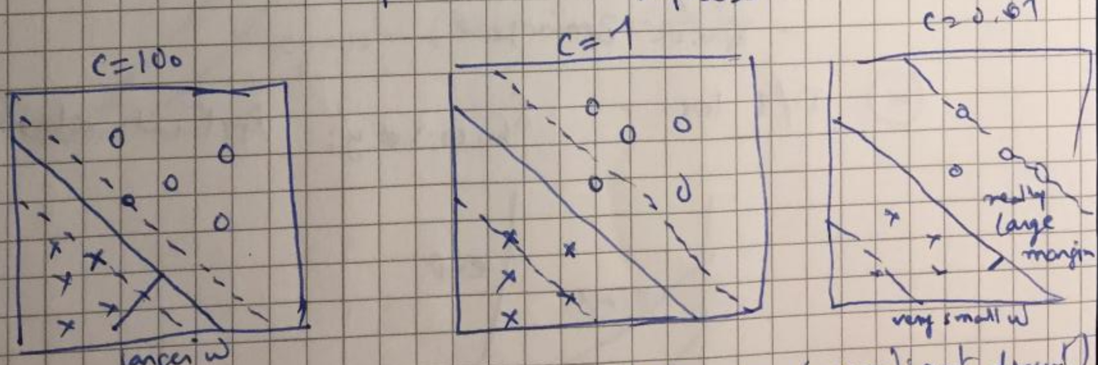
soft SVM
$$\begin{cases} y_i(w^T x_i + b) \geq 1 - \xi_i \quad \forall_i \, \xi_i \geq 0 \\ \text{and } \min_{w,b} \; w^T w + C \sum_{i=1}^{n} \xi_i \end{cases}$$

Useful when mislabelled data. with a small c $10^{-4}, 10^{2}$

$$\xi_i = \begin{cases} 1 - y_i(w^T x_i + b) & \text{if } y_i(w^T x_i + b) < 1 \\ 0 & \text{if } y_i(w^T x_i + b) \geq 1 \end{cases}$$

OR, $\xi_i = \max(1 - y_i(w^T x_i + b), 0)$

So now,
$$\min_{w,b} \; \underbrace{w^T w}_{\text{Regularization}} + \underbrace{C \sum_i \max(1 - y_i(w^T x_i + b), 0)}_{\text{hinge loss}}$$



soft-SVM formulation like logistic regression (gradient descent) with hinge loss instead of logistic loss.

Generally in ML:

$$\min \frac{1}{n} \underbrace{\sum_i l(h_w(x_i, y_i))}_{\text{loss}} + \underbrace{\lambda r(w)}_{\text{Regularization}}$$

loss
Improves prediction

Regularization
reduces complexity of model

Common loss functions:

① Hinge loss $\max(1 - h_w(x) y_i, 0]^p$

E.g. $h_w(x) = w^T x + b$ for linear classifier

$p = 1$ SVM
$p = 2$ Squared loss SVM, closed form sol?

② Log loss $\log(1 + e^{-y_i h_w(x_i)})$ Logistic Regression

Can get well calibrated probabilities. Eg.

$$p(y|x) = \frac{1}{1 + e^{-y_i h(w)}}$$

③ Exponential loss $e^{-y_i h_w(x_i)}$ Adaboost

- Any tiniest mistakes, freaks out, just to get one outlier right.
- Noisy data problem
- Quick converges

④ 0/1 loss - $\delta_{h_w(x_i) \neq y_i}$ Actual classification loss.