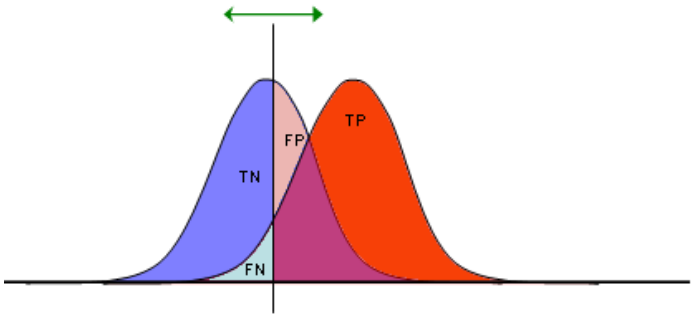
Problem Set 3: KRR, CV

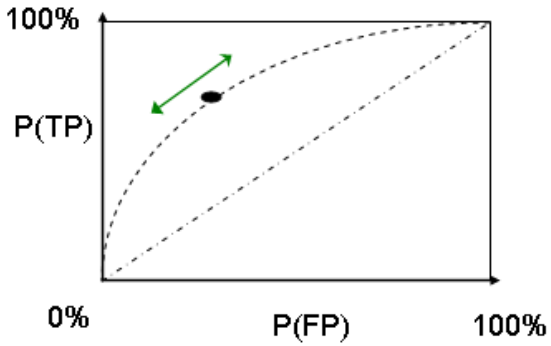
Benjamin Pietrowicz, Budi Yanto

June 25, 2014

ROC Characteristic



TP	FP
FN	TN
1	1



Source: `http://en.wikipedia.org/wiki/Receiver_operating_characteristic`

TPR, FPR

$$TN(z) = \Phi(z - \mu_N) \qquad FP(z) = 1 - TN(z) = \Phi(\mu_n - z)$$

$$FN(z) = \Phi(z - \mu_P) \qquad TP(z) = 1 - FN(z) = \Phi(\mu_p - z)$$

$$TPR(z) = \frac{TP(z)}{TP(z) + FN(z)} \qquad FPR(z) = \frac{FP(z)}{FP(z) + TN(z)}$$

TPR, FPR

$$TN(z) = \Phi(z - \mu_N) \qquad FP(z) = 1 - TN(z) = \Phi(\mu_n - z)$$

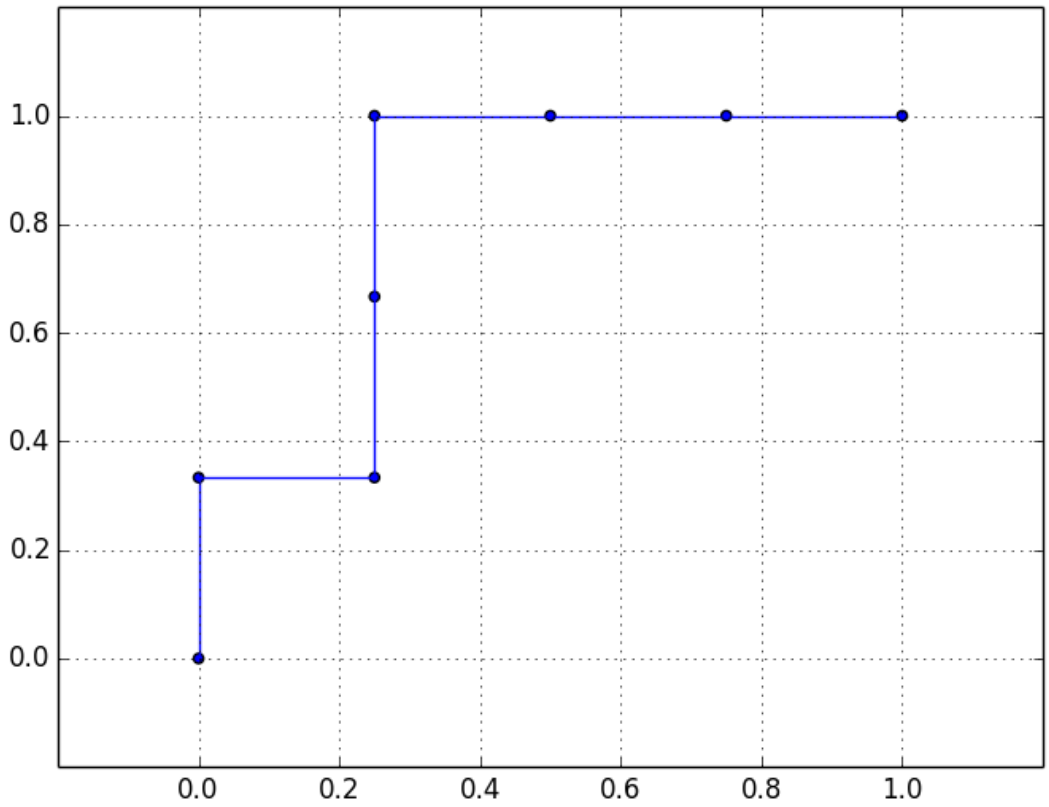
$$FN(z) = \Phi(z - \mu_P) \qquad TP(z) = 1 - FN(z) = \Phi(\mu_p - z)$$

$$TPR(z) = \frac{TP(z)}{TP(z) + FN(z)} \qquad FPR(z) = \frac{FP(z)}{FP(z) + TN(z)}$$

Unconditional distribution:

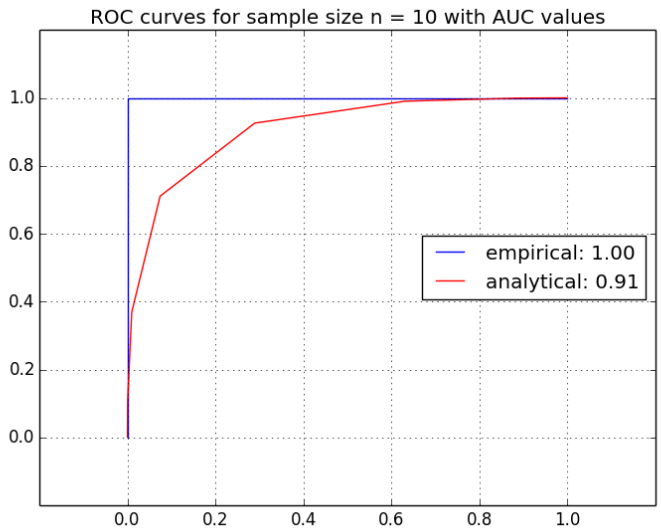
$$p(x) = p(x|y = -1) \cdot p(y = -1) + p(x|y = +1) \cdot p(y = +1)$$

ROC Curves - 1D-Toy Data Set



sorted samples: $-1|+1|-1|-1|+1|+1|+1$

ROC Curves - 1D-Toy Data Set



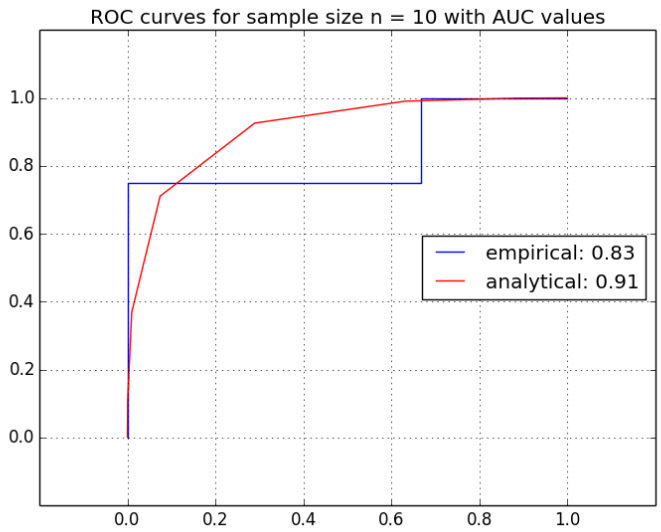
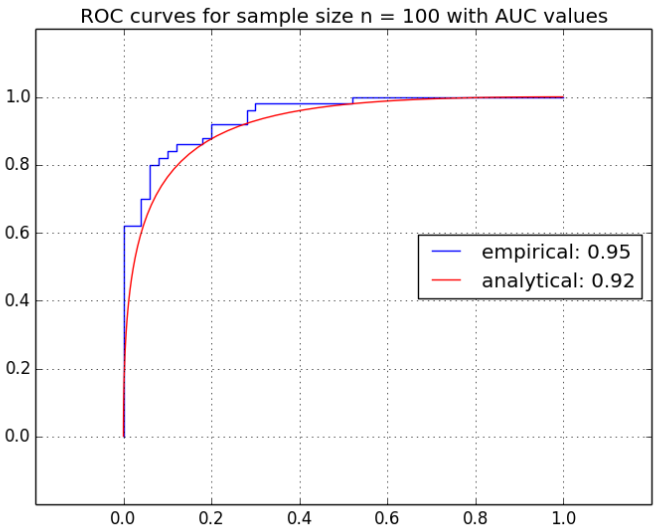


Figure : ROC curves for sample size n = 10

ROC Curves - 1D-Toy Data Set



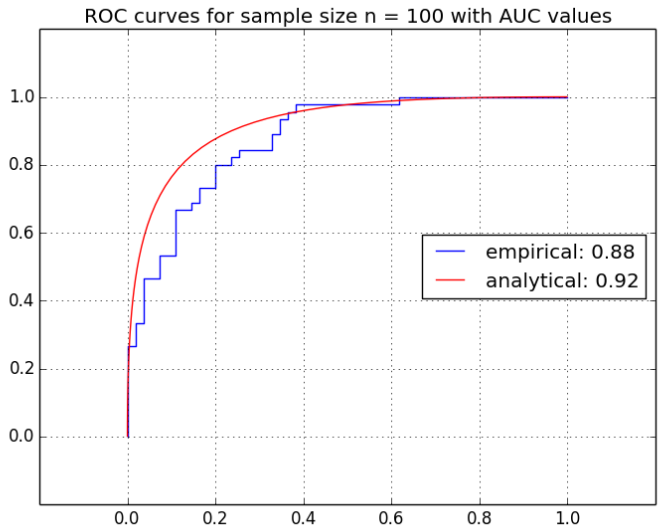
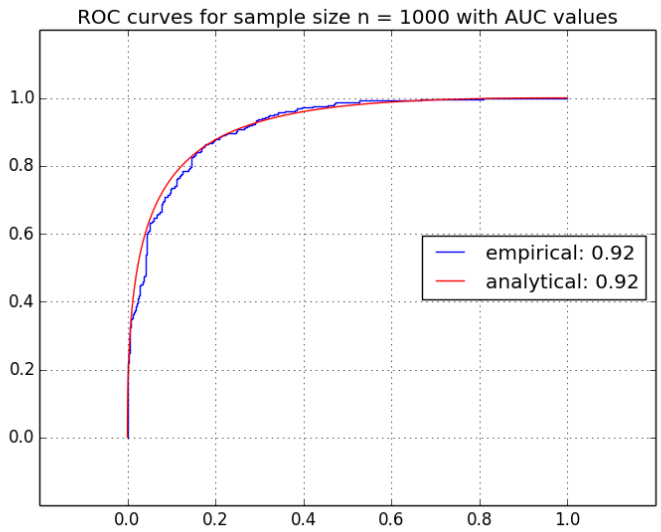


Figure : ROC curves for sample size n = 100

ROC Curves - 1D-Toy Data Set



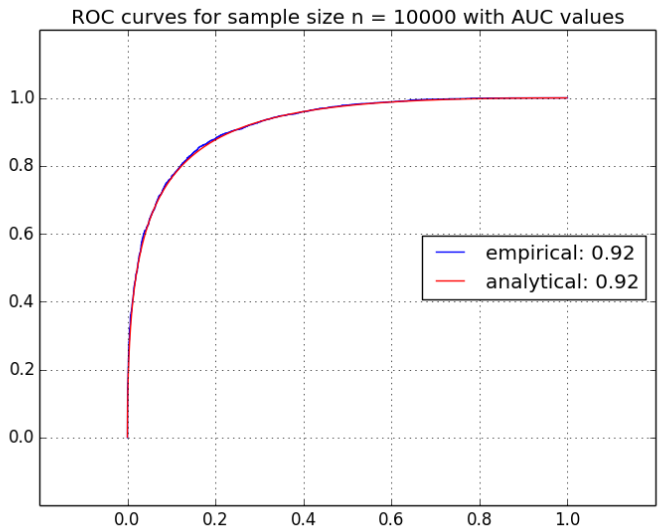


Figure : ROC curves for sample size n = 1000 and n = 10000

Data Sets

Data Set	Dimension	Total Training Data	Total Test Data
Banana	2	400	4900
Diabetis	8	468	300
Flare-Solar	9	666	400
Image	18	1300	1010
Ringnorm	20	400	7000

Cross-Validation Parameters

- **Cross-Validation** with *nrepetitions* = 2 and *nfolds* = 10
- **LOOCV**: regularization = [0]

Kernel	**Kernel Parameters**	**Regularization**
Linear	[0]	np.logspace(-2,2,10)
Polynomial	np.arange(1,10)	np.logspace(-2,2,10)
Gaussian	np.logspace(-2,2,10)	np.logspace(-2,2,10)

Cross-Validation Results

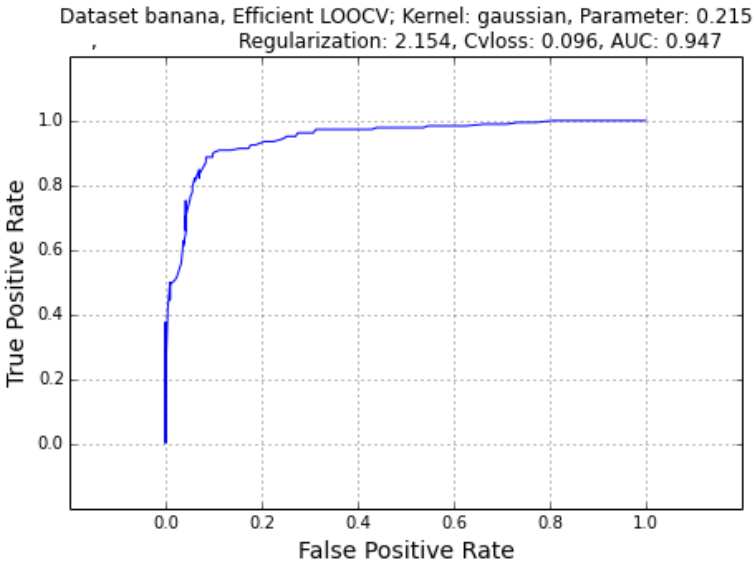
Data Set	Kernel	Parameter	Regularization	Cvloss
Banana	Gaussian	0.215	2.154	0.096
Diabetis	Gaussian	4.642	2.154	0.237
Flare-Solar	Polynomial	2.0	762.223	0.347
Image	Gaussian	0.599	0.464	0.028
Ringnorm	Gaussian	4.642	0.464	0.062

Table : LOOCV

Data Set	Kernel	Parameter	Regularization	Cvloss
Banana	Gaussian	0.215	4.642	0.096
Diabetis	Gaussian	35.938	0.010	0.241
Flare-Solar	Gaussian	35.938	0.010	0.348
Image	Gaussian	1.668	0.028	0.022
Ringnorm	Gaussian	4.642	0.077	0.039

Table : Standard CV

ROC Curves - Banana



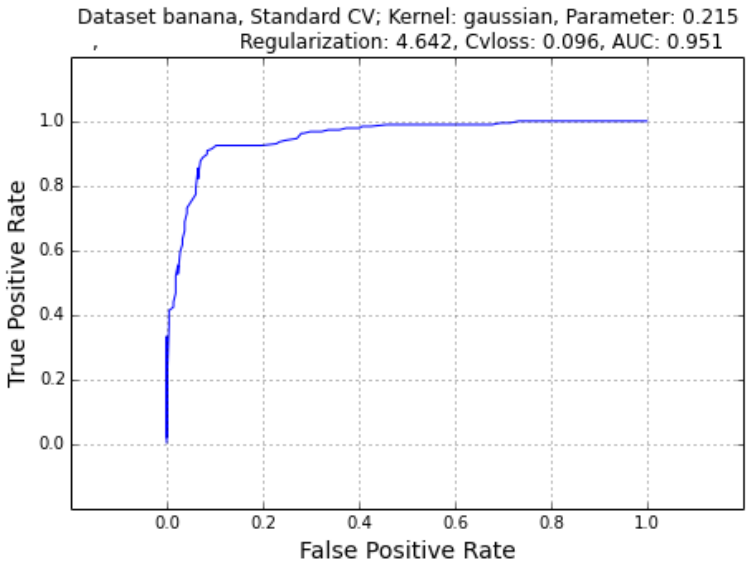
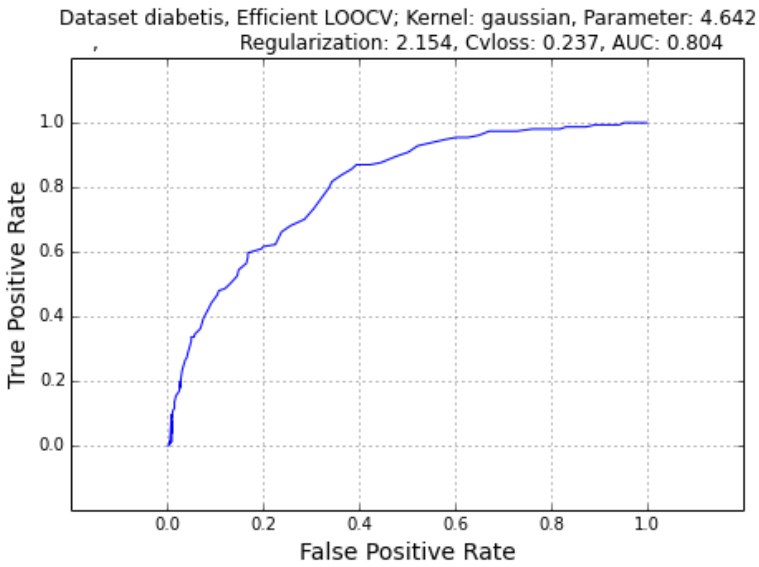


Figure : ROC curves of banana dataset

ROC Curves - Diabetis



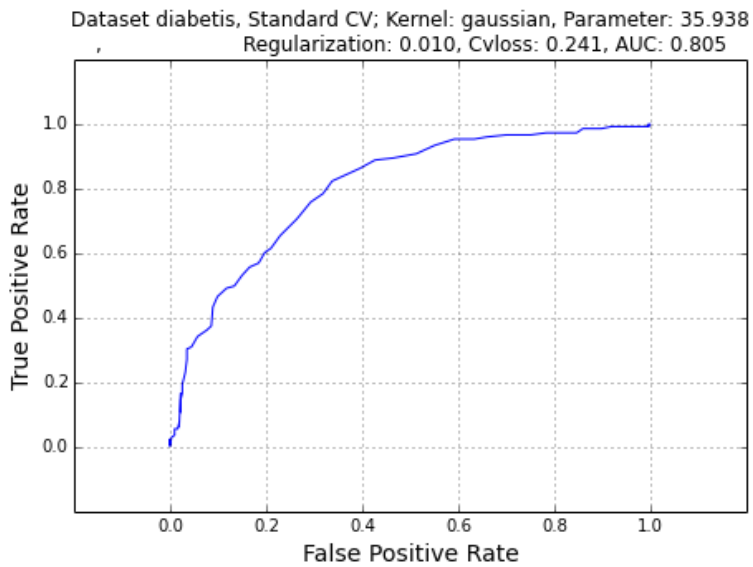


Figure : ROC curves of diabetis dataset

ROC Curves - Flare-Solar

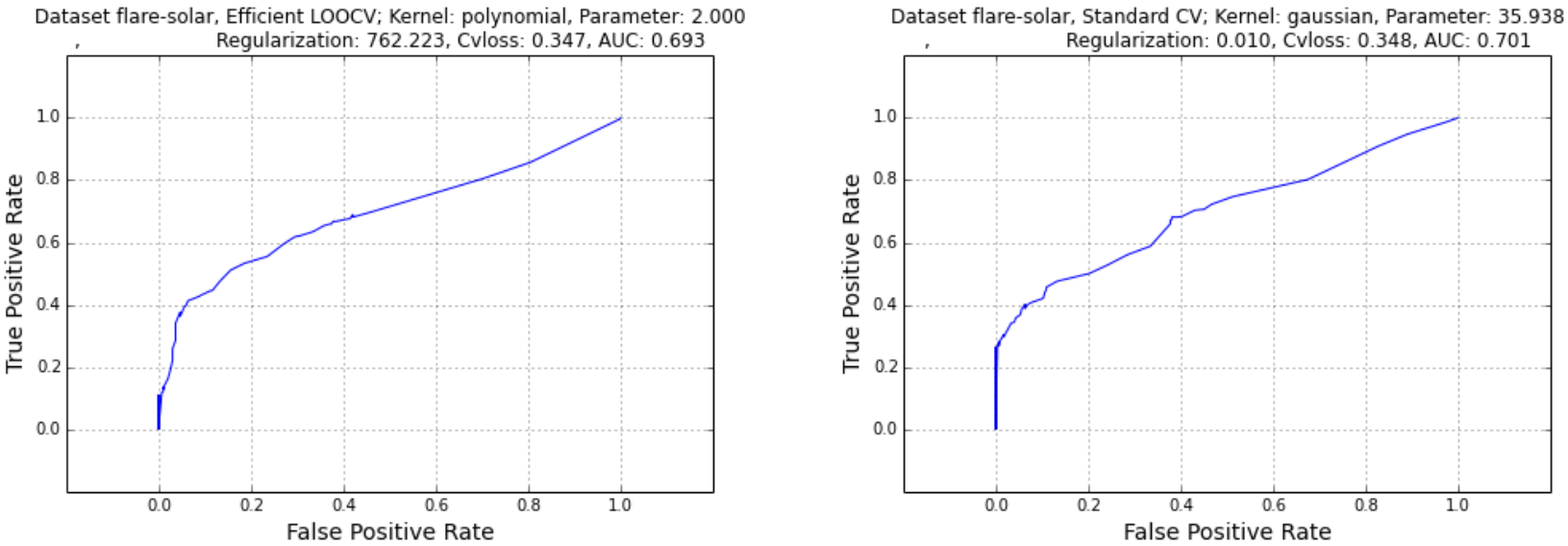
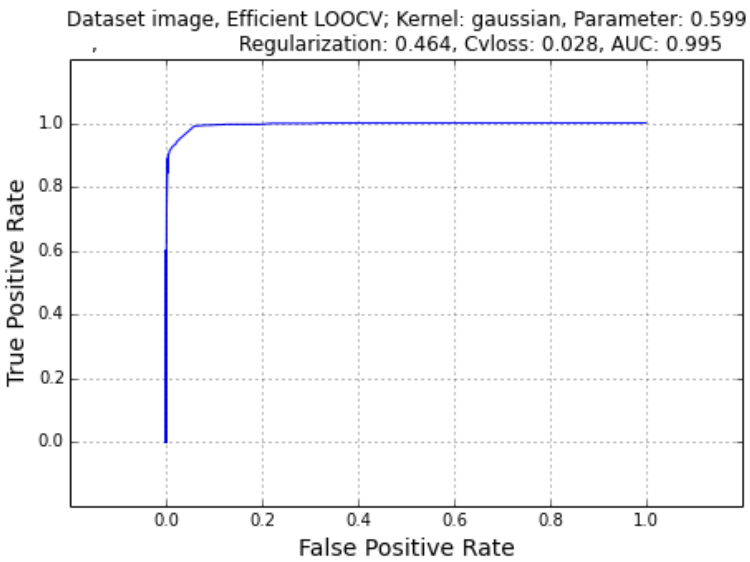


Figure : ROC curves of flare-solar dataset

ROC Curves - Image



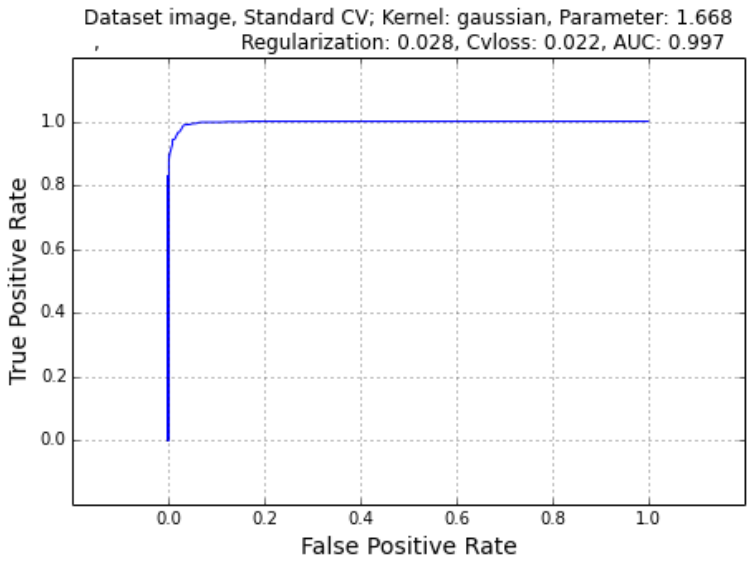
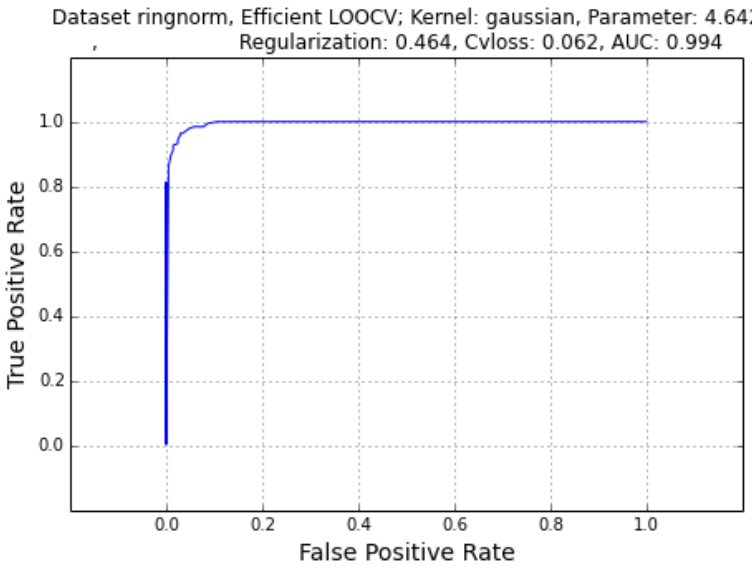


Figure : ROC curves of image dataset

ROC Curves - Ringnorm



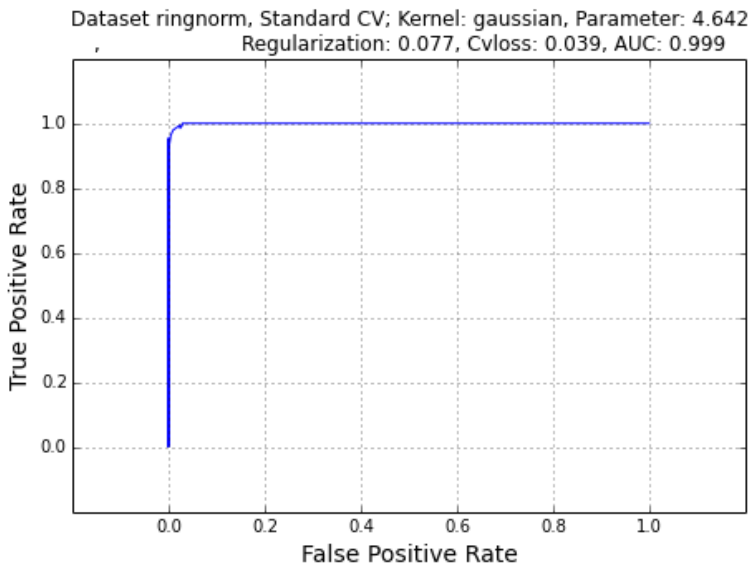


Figure : ROC curves of ringnorm dataset

Correspondence between Cvloss and AUC

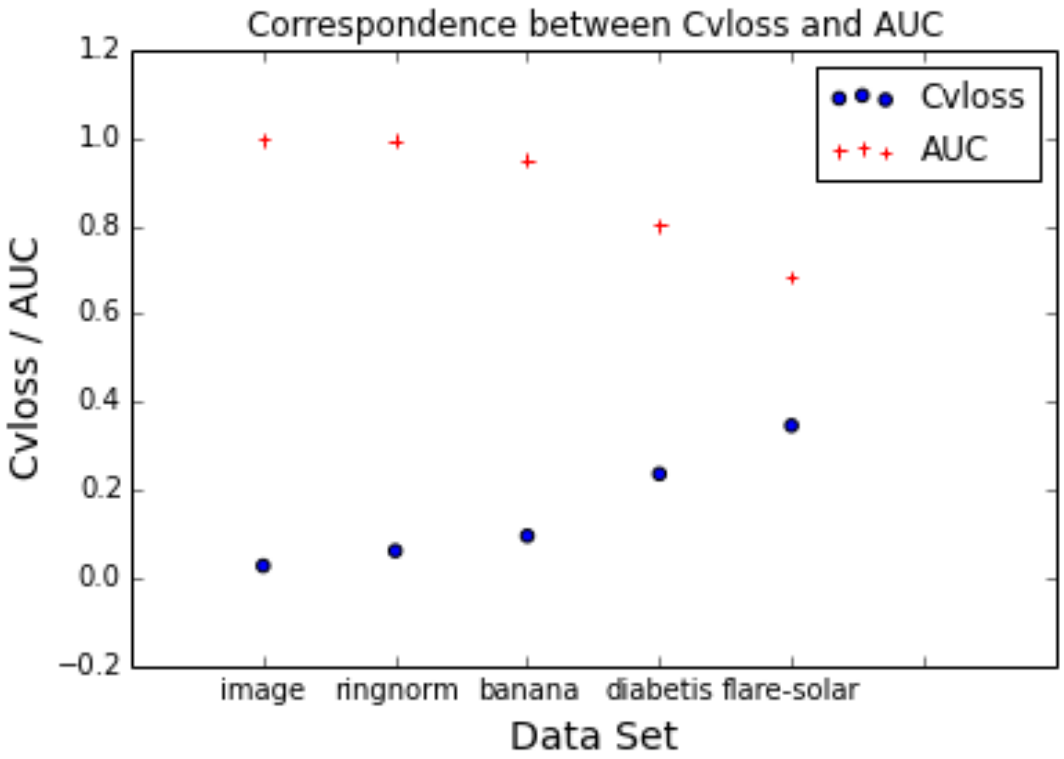


Figure : Correspondence between Cvloss and AUC

Comparison of Standard CV and LOOCV

Data Set	Cvloss LOOCV	Cvloss St. CV
Banana	0.096	0.096
Diabetis	0.237	0.241
Flare-Solar	0.347	0.348
Image	0.028	0.022
Ringnorm	0.062	0.039

Data Set	Time LOOCV (in sec.)	Time St. CV (in sec.)
Banana	109.220	78.326
Diabetis	237.252	113.388
Flare-Solar	466.499	217.701
Image	4596.275	1420.264
Ringnorm	126.286	96.347

Thank you !!!