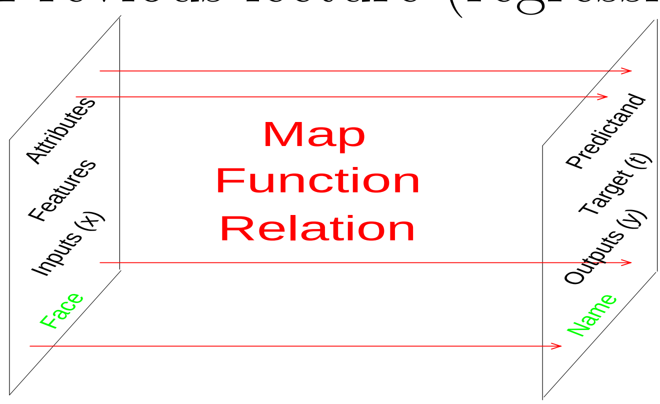
CLASSIFICATION (DISCRIMINATION)

Caren Marzban

http://www.nhn.ou.edu/~marzban

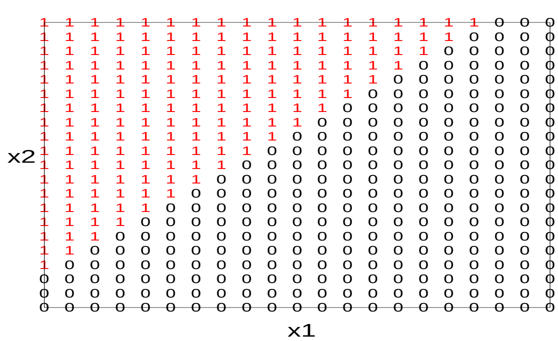
Generalities

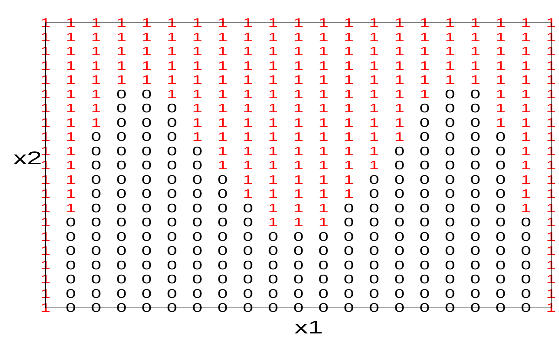
Prerequisite: Previous lecture (regression NN)



Regression: Continuous target.

Classification: Discrete target (e.g., target$= 0, 1$).





Discriminant Analysis

Given one input variable, x, what's the class (i=0,1 - nontornado/tornado)?
Gaussian distribution.

$$L_i(x) \sim e^{-\frac{(x-\mu_i)^2}{2\sigma_i^2}} \sim prob(x|C=i).$$

$$P_i(x) = \frac{p_i L_i(x)}{p_0 L_0(x) + p_1 L_1(x)} \sim prob(C=i|x),$$

$$p_i = \text{prior (climatological) prob.} = N_i/N.$$

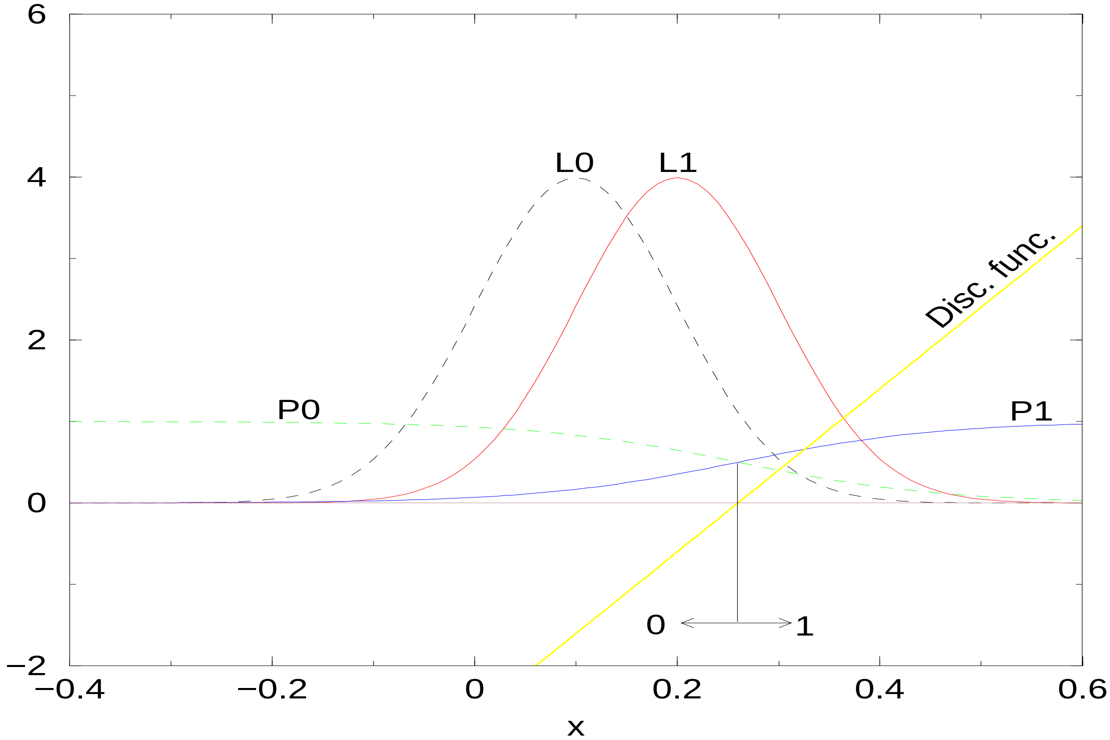
$$(L_0(x){+}L_1(x) \neq 1,\ \ p_0{+}p_1 = 1,\ \ P_0(x){+}P_1(x) = 1)$$

If $P_1(x) > P_0(x)$, then $x \in 1$.

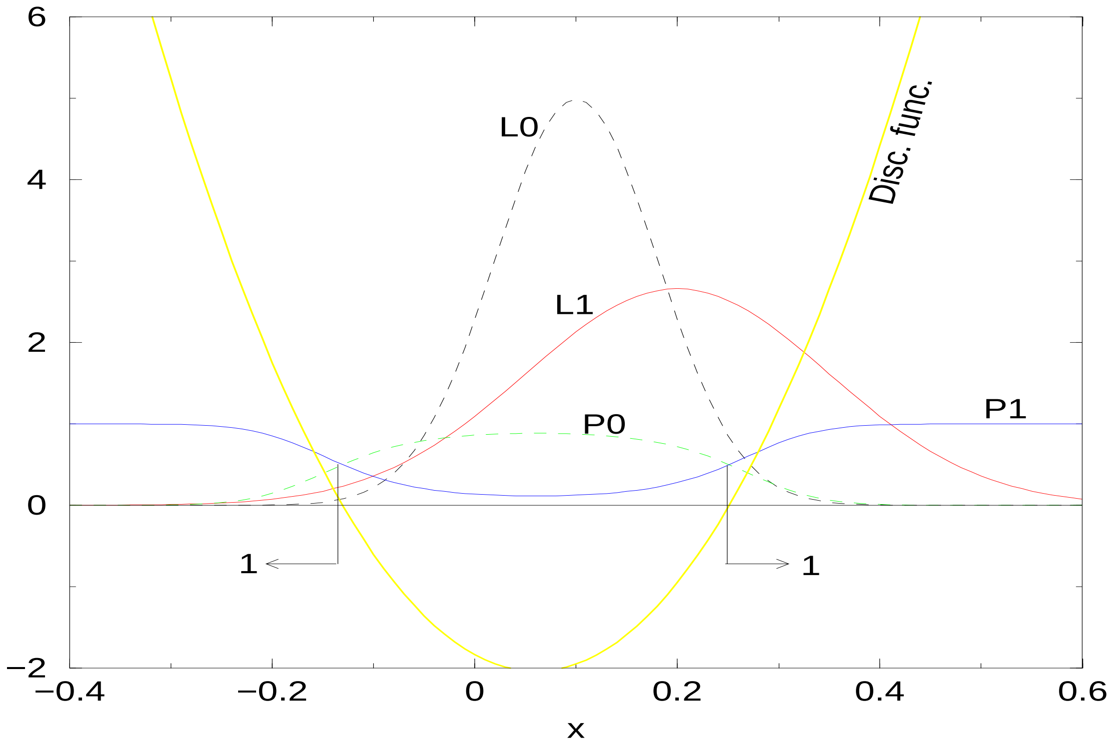
$$\log\ (P_1(x)/P_0(x)) \equiv \text{Discriminant function}$$
$$\sim\ (\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2})x^2 - 2(\frac{\mu_0}{\sigma_0^2} - \frac{\mu_1}{\sigma_1^2})x + \text{constant}\ \ .$$

Gaussian $\longrightarrow$ quadratic discrim.
Gaussian + equivariant ($\sigma_0 = \sigma_1$) $\longrightarrow$ linear discrim.
(Recall: y =linear, E=MSE, errors=gaussian)

Linear Discriminant



Quadratic Discriminant



NN

From previous section

$$P_1(x) = \frac{1}{1+e^{-(\text{discrim. func.})}}.$$

If discrim. func. = linear, then $P_1(x) =$ logistic.
I.e., NN with no hidden nodes.

$$y(x,\omega) = g\left(\sum_{i=1}^{H} \omega_i \, f(\sum_{j=1}^{N_{in}} \omega_{ij} x_j - \theta_j) - \omega\right)$$

But while $t = (0,1)$, $y \neq (0,1)$.

Minimizing cross entropy

$$E = -\frac{1}{N} \Sigma [t \log y(x,\omega) + (1-t) \log(1 - y(x,\omega))] \ \ ,$$

guarantees

$$y(x,\omega) = P_1(x) = prob(C = 1|x) \ \ ,$$

if f and g logistic.

$$\boxed{\text{E=cross-entropy \& f=g=logistic} \Rightarrow \text{output=prob}}$$

Analogous to

$$\text{E=MSE} \Rightarrow y(x,\omega) = < t|x >$$

for regression NN (See Appendix A, previous lecture).

Method

Mostly, same as regression:
Weight-decay
Local minima
Bootstrap ...

Basically, the seed-trial table (previous lecture).

One difference:
E=cross-entropy=unique, if output=post. prob.

One more difference: Rare events ($p_1 << p_0$).
Weigh cross-entropy according to N_0 and N_1.
Artificial balancing.
Rescale output to recover posterior probability:

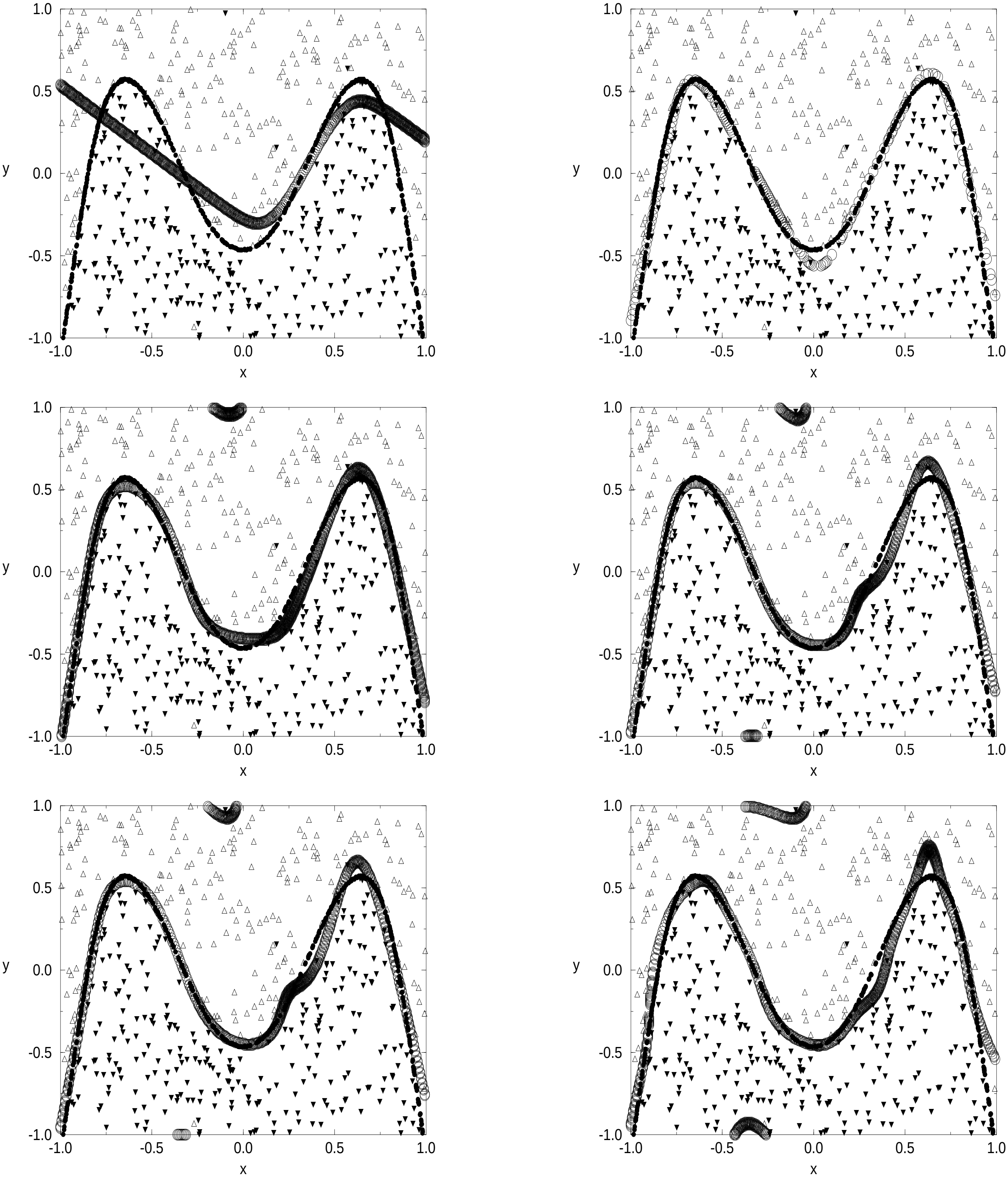
$$new_1 = \frac{(\frac{p_1}{pd_1})\ old_1}{(\frac{p_0}{pd_0})\ old_0 + (\frac{p_1}{pd_1})\ old_1},$$

where p_i = true climatological prob, pd_i = the one used in trn; $p_0 + p_1 = pd_0 + pd_1 = out_0 + out_1 = 1$.

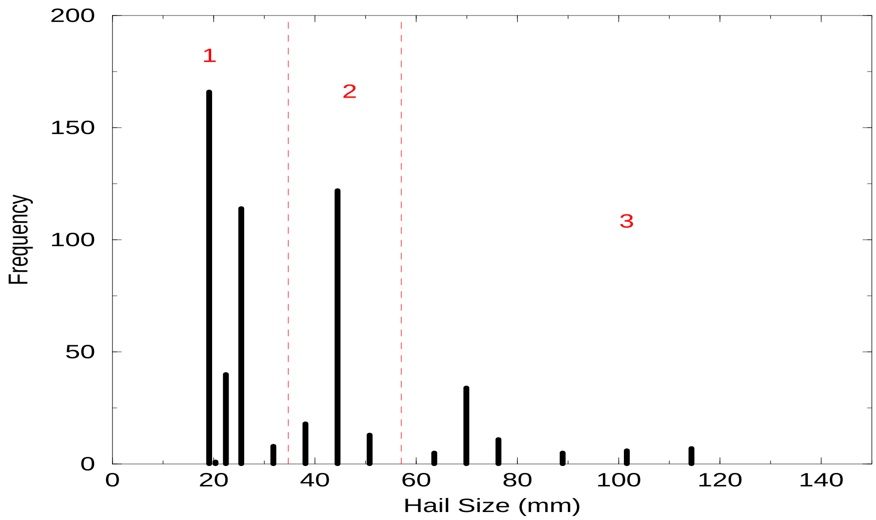
One final difference: Performance.
probabilistic forecast $\rightarrow$ probabilistic framework.

Overfitting

Recall overfitting from previous lecture (H=2,4,6,8,10,15).



Practical Application - Hail Size

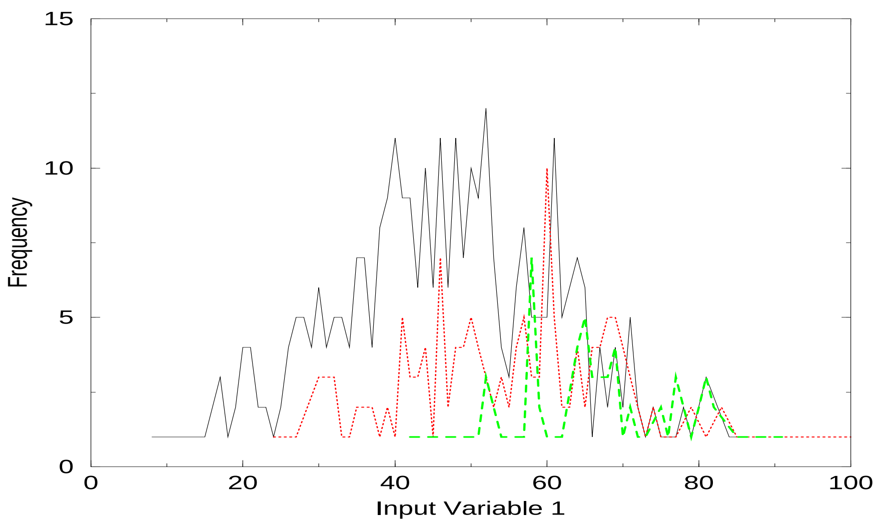


The distribution of hail-size. Note 3 classes.

Regression NN $\rightarrow$ hail size in mm.
Classification NN $\rightarrow$ prob. of hail size class (1,2,3).

Preprocess like regression NN, but condition on class.

Check class-conditional input distribution/histogram.



Exclude collinear inputs, based on class-conditional r.

Etc.

Performance

$$N_{\text{in}} = 9 \ \ , H_{\text{optimal}} = 4$$

$$N_{\text{out}} = 3 \ \ \text{(one per class)}$$

Target encoding: (1,0,0)=1, (0,1,0)=2, (0,0,1)=3.

Categorical: Threshold probabilities.

$$\text{C-table} = \begin{pmatrix} C_{11} & C_{12} & C_{13} \\ C_{21} & C_{22} & C_{23} \\ C_{31} & C_{32} & C_{33} \end{pmatrix},$$

C_{12} = No. class 1 cases (incorrectly) classified in 2.

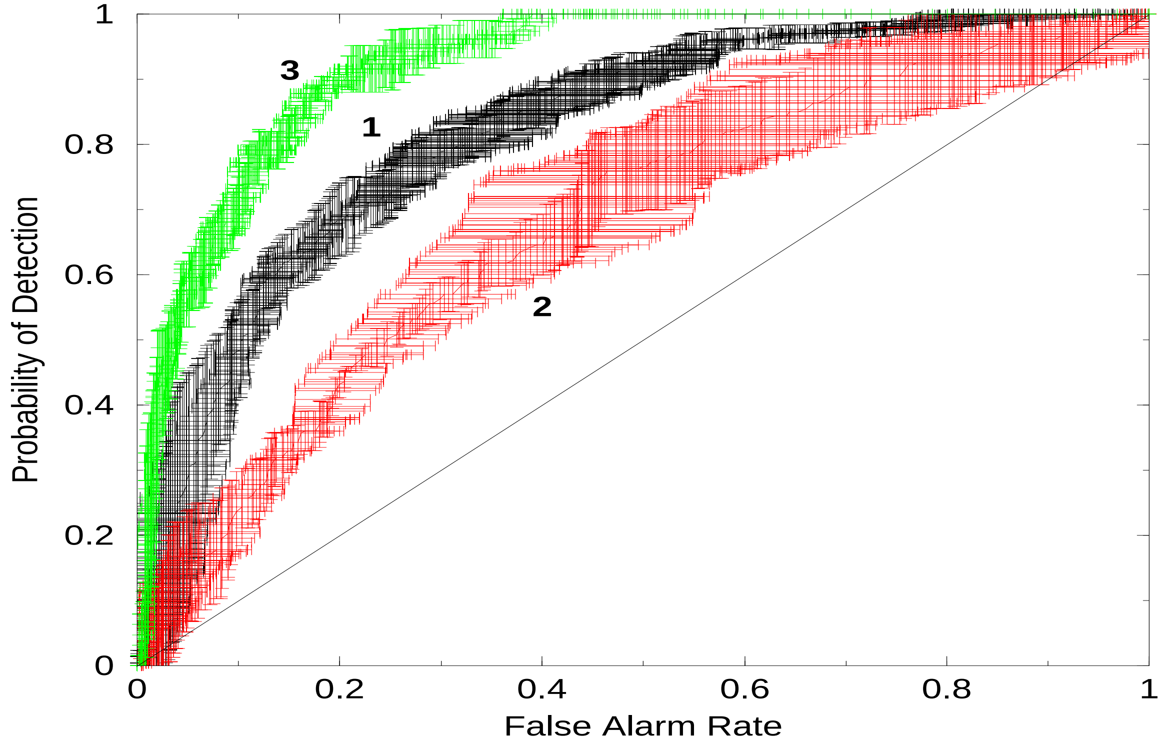
3-class problem $\longrightarrow$ three 2-class problems:
1 or not, 2 or not, 3 or not.

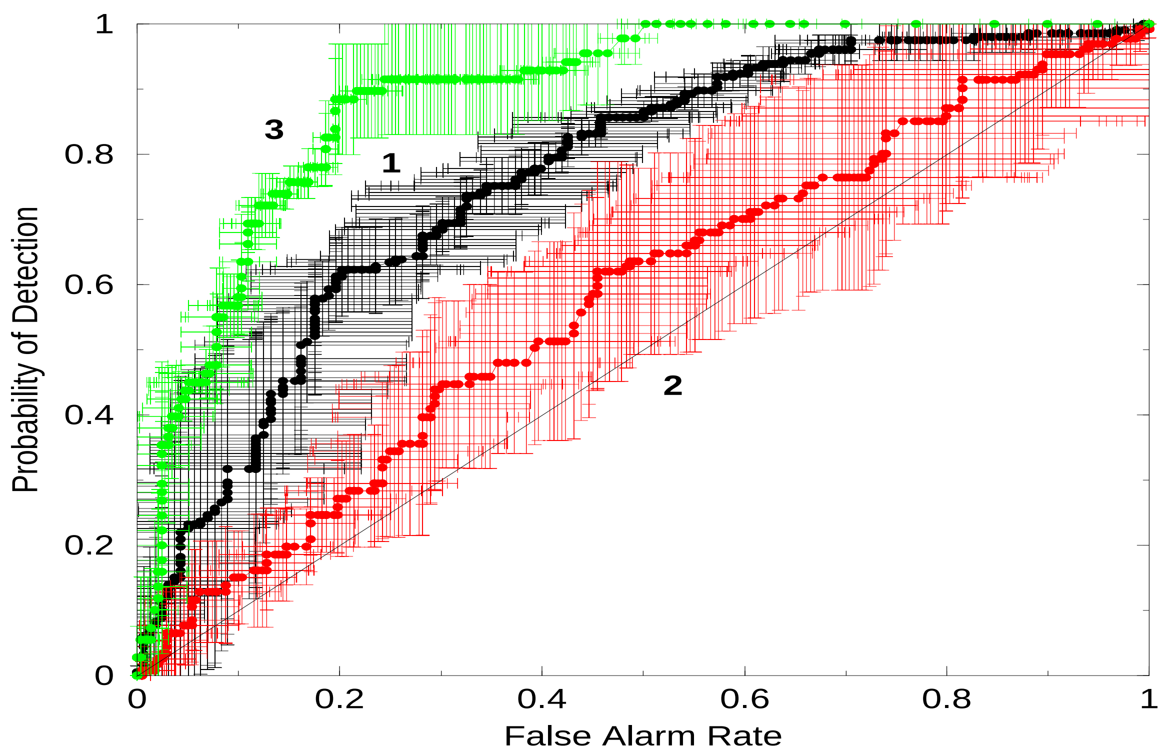
$$\text{C-table} = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix}.$$

Probability of Detection and False Alarm *Rate*:

$$\text{POD} = \frac{C_{22}}{C_{21} + C_{22}}, \qquad \text{FAR} = \frac{C_{12}}{C_{11} + C_{12}}.$$

ROC Diagram: POD vs. FAR as threshold $0 \rightarrow 1$.
Trn set (top), and vld set (bottom):





Probabilistic: (see lecture notes)

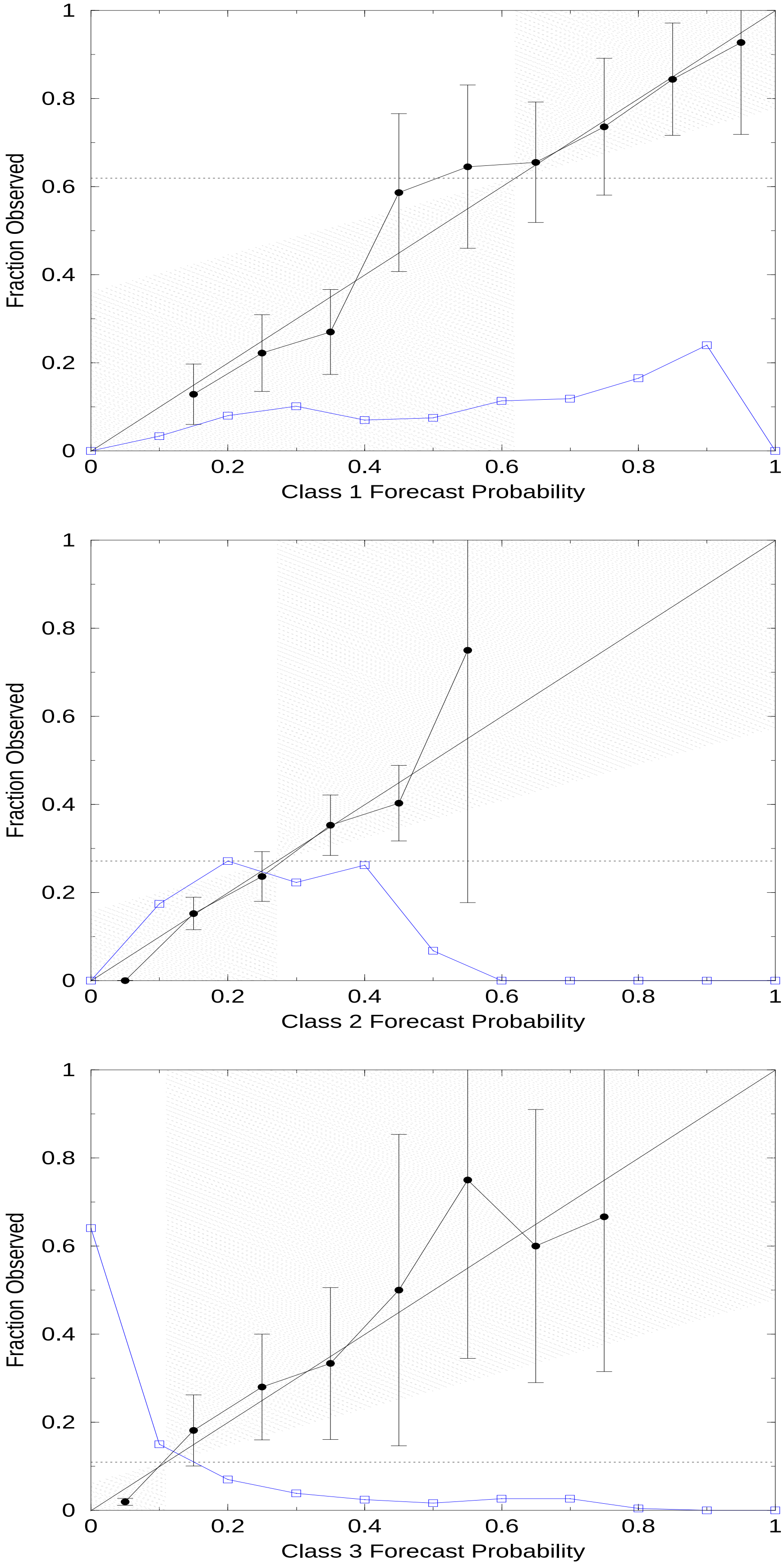
3 concepts, 3 questions:

- Reliability: Do the forecast probs match observed frequency?

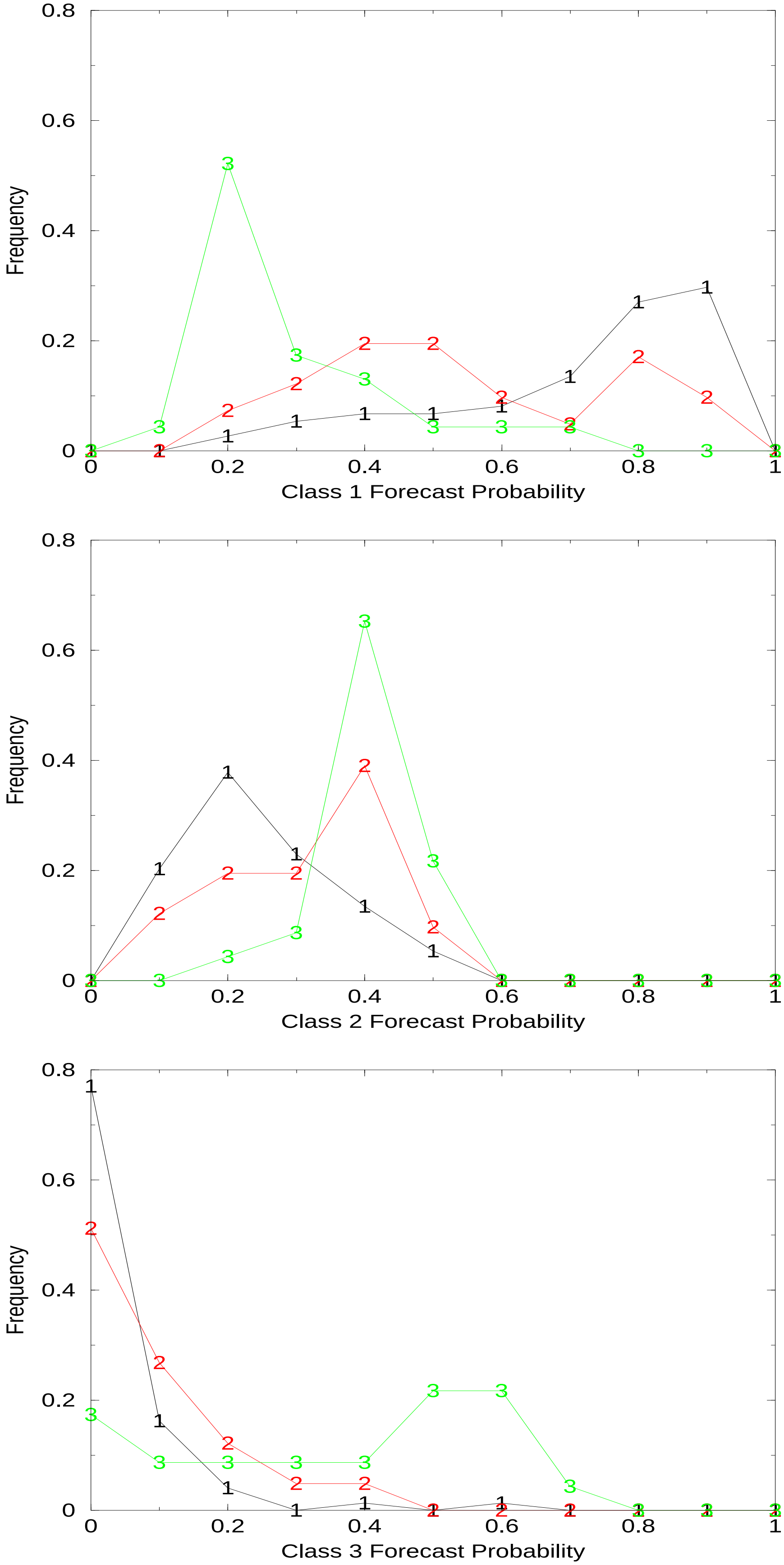
- Refinement: What's the distribution of the forecasts?

- Discrimination: What's the conditional distribution of the forecasts?

  More on these in my verification talk, later.



Fraction Observed
Class 1 Forecast Probability
Fraction Observed
Class 2 Forecast Probability
Fraction Observed
Class 3 Forecast Probability



Frequency
Class 1 Forecast Probability
Frequency
Class 2 Forecast Probability
Frequency
Class 3 Forecast Probability

Conclusion

We are done.

We did regression and classification.

We developed a Hail NN. Output:
Forecast hail size = 24mm
Prob. of coin-size hail = 80%
Prob. of golfball-size hail = 15%
Prob. of baseball-size hail = 5%

We learned performance assessment.

Verification of probabilistic forecasts. E.g.,
Forecast probabilities are highly reliable, reasonably refined, and quite discriminatory.

Better not quantify too much.

Single Lesson

NNs do statistics. So, start from discriminant analysis, or logistic regression.