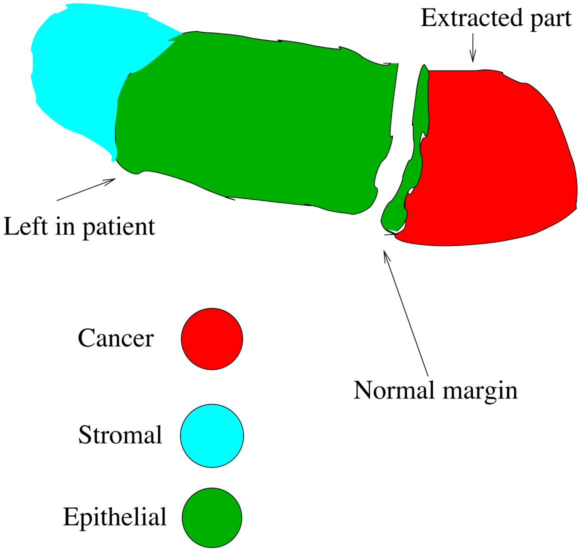
Cancer detection via the lasso and customized training

Robert Tibshirani, Stanford University

Google $->$ Tibshirani
(WARNING: top link might be that imposter)

Machine learning department, CMU; 2014


Extracted part
Left in patient
Normal margin
Cancer
Stromal
Epithelial

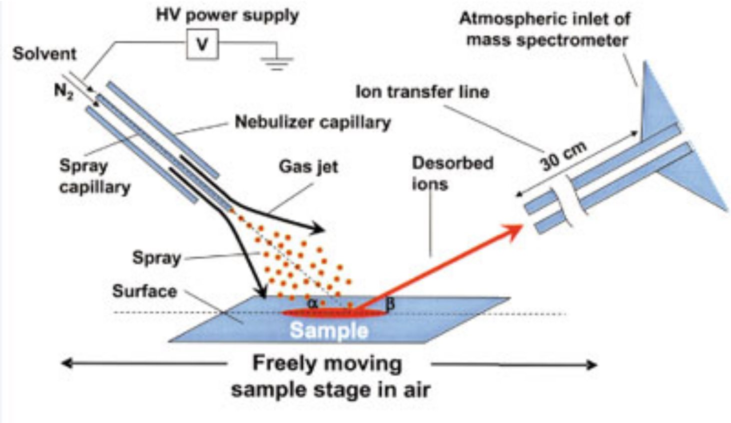
The challenge

- Build a classifier than can distinguish three kinds of tissue: normal epithelial, normal stromal and cancer.
- Such a classifier could be used to assist surgeons in determining, in real time, whether they had successfully removed all of the tumor. Current pathologist error rate for the real-time call can be as high as 20%.

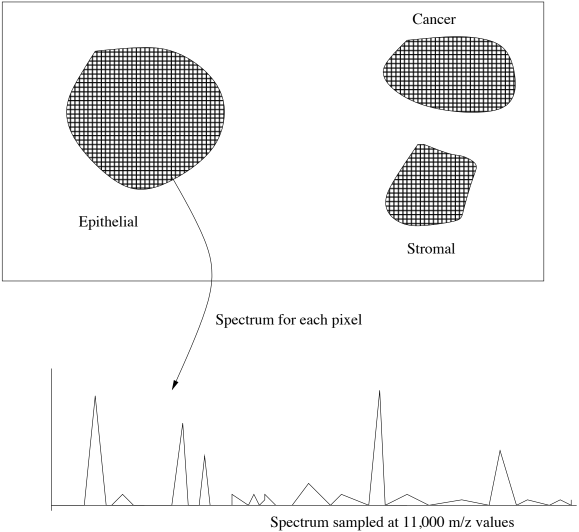
Technology to the rescue!

DESI (Desorption electrospray ionization)

An electrically charged “mist” is directed at the sample; surface ions are freed and enter the mass spec.



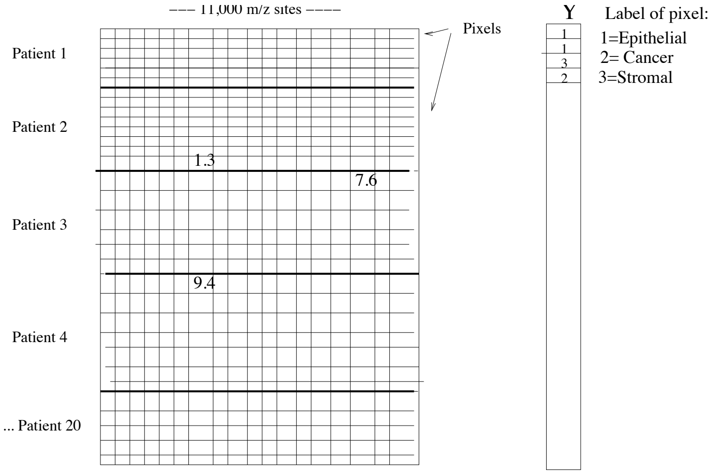
The data for one patient



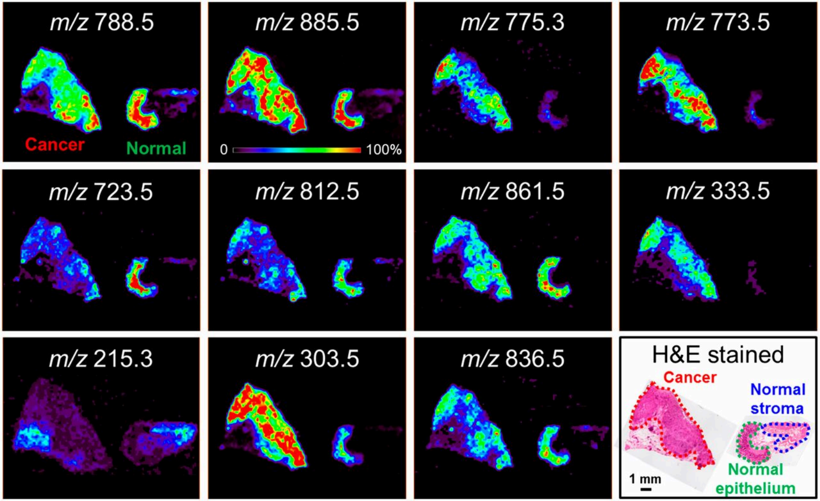
Details

- 20 patients, each contributing a sample of epithelial, stromal and cancer tissue.
- Labels determined after 2 weeks of testing in pathology lab.
- At each pixel in the image, the intensity of metabolites is measured by DESI. Peaks in the spectrum representing different metabolites.
- The spectrum has been finely sampled, with the intensity measured at about $11,000$ m/z sites across the spectrum, for each of about 8000 pixels.

The overall data



Selected negative ion mode DESI-MS ion images of sample GC727.



What we need to tackle this problem

- A statistical classifier (algorithm) that sorts through the large number of features, and finds the most informative ones: a sparse set of features.
- We are doing pixel-wise classification

The Lasso

- Regression problem: We observe n feature-response pairs $(x_i,\ y_i)$, where x_i is a p-vector and y_i is real-valued.
- Let $x_i = (x_{i1}, x_{i2}, \ldots x_{ip})$
- Consider a *linear regression model*:

$$y_i = \beta_0 + \sum_j x_{ij}\beta_j + \epsilon_i$$

  where ϵ_i is an error term with mean zero. β_j is the weight given feature j
  Later: y_i will take one of 3 values (epithelial, stromal, cancer) and x_{ij} will be the height of the spectrum for patient i, at m/z site j.
- Least squares fitting is defined by

$$\underset{\beta_0,\beta}{\text{minimize}} \frac{1}{2}\sum_i (y_i - \beta_0 - \sum_j x_{ij}\beta_j)^2$$

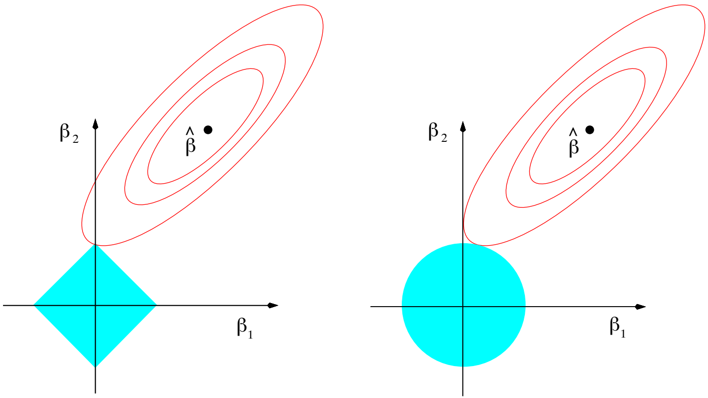
The Lasso— continued

The **Lasso** is an estimator defined by the following optimization problem:

$$\underset{\beta_0,\beta}{\text{minimize}} \frac{1}{2}\sum_i (y_i - \beta_0 - \sum_j x_{ij}\beta_j)^2 \qquad \text{subject to} \quad \sum |\beta_j| \leq s$$

- Penalty $\implies$ sparsity (feature selection)
- Convex problem (good for computation and theory)
- *Ridge regression* uses penalty $\sum_j \beta_j^2 \leq s$ and does not yield sparsity

Why does the lasso give a sparse solution?



Lasso $\sum_j |\beta_j| \leq s$

Ridge $\sum_j \beta_j^2 \leq s$

Back to our problem

- $K = 3$ classes (epithelial, stromal, cancer): multinomial model

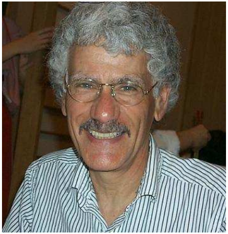
$$\log \frac{Pr(Y_i = k|x)}{\sum_k Pr(Y_i = k|x)} = \beta_{0k} + \sum_j x_{ij}\beta_{jk},\ k = 1, 2, \ldots K$$

  Here x_{ij} is height of spectrum for sample i at jth m/z position

- We replace the least squares objective function by the multinomial log-likelihood
- Add lasso penalty $\sum |\beta_j| \leq s$; optimize, using cross-validation to estimate best value for budget s.
- yields a pixel classifier, and also reveals which m/z sites are informative.

Fast computation is essential

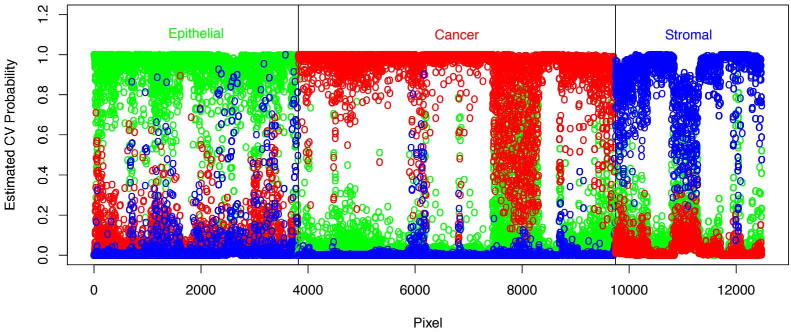
- Our lab has written a open-source R language package called `glmnet` for fitting lasso models. Written in FORTRAN!!!!!
- It is very fast- can solve the current problem in a few minutes on a PC. Some builtin parallelization too.
- Not "off-the shelf": Many clever computational tricks were used to achieve the impressive speed.
- Lots of features- Gaussian, Logistic, Poisson, Survival models; elastic net; grouping; parameter constraints; Available in R and Matlab.

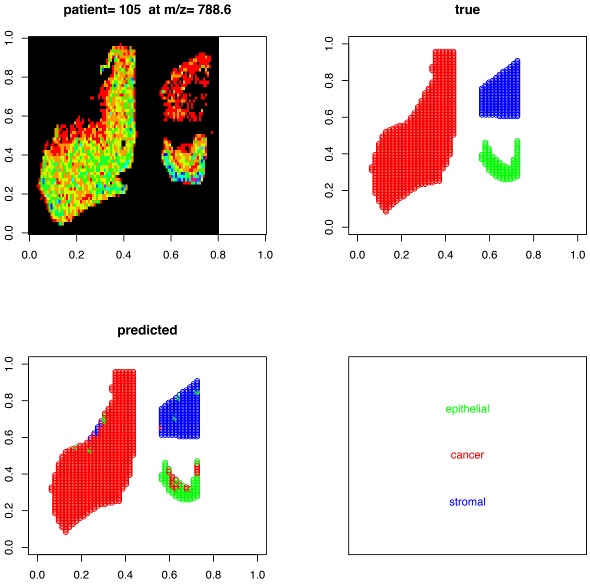


Jerry Friedman

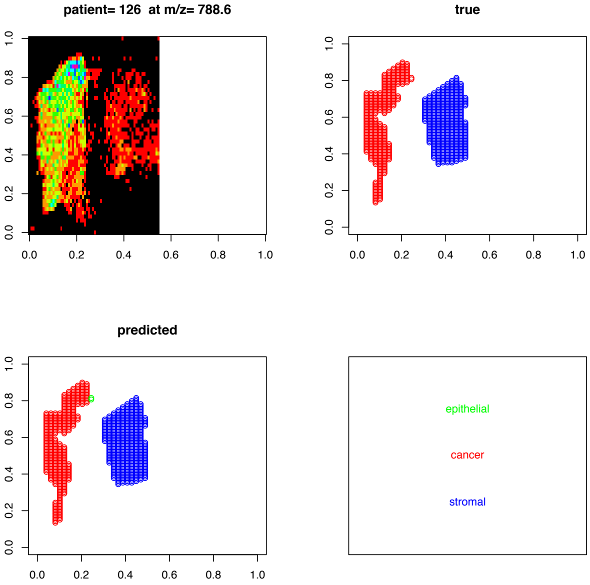
Trevor Hastie

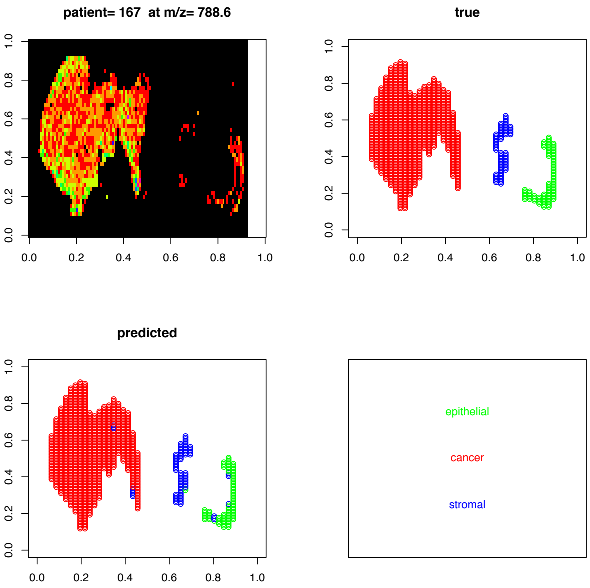
Cross-validated estimates of class probabilities





patient= 105 at m/z= 788.6
true
predicted
epithelial
cancer
stromal
0.0
0.2
0.4
0.6
0.8
1.0


patient= 126 at m/z= 788.6
true
predicted
epithelial
cancer
stromal
0.0
0.2
0.4
0.6
0.8
1.0


patient= 167 at m/z= 788.6
true
predicted
epithelial
cancer
stromal
0.0
0.2
0.4
0.6
0.8
1.0

Other approaches

- Support vector machines: classification error was a little higher than lasso; doesn't give a sparse solution easily
- Deep learning (with help from a student of Geoff Hinton): reported that it didn't work any better than lasso; thought that non-linearities were likely unimportant for this problem, and sparsity was more important