FIGURE 18.1. Test-error results for simulation experiments. Shown are boxplots of the relative test errors over 100 simulations, for three different values of $p$, the number of features. The relative error is the test error divided by the Bayes error, $\sigma^2$. From left to right, results are shown for ridge regression with three different values of the regularization parameter $\lambda$: 0.001, 100 and 1000. The (average) effective degrees of freedom in the fit is indicated below each plot.
FIGURE 18.2. Soft thresholding function \( \text{sign}(x)(|x| - \Delta)_+ \) is shown in orange, along with the 45° line in red.
FIGURE 18.3. Heat-map of the chosen 43 genes. Within each of the horizontal partitions, we have ordered the genes by hierarchical clustering, and similarly for the samples within each vertical partition. Yellow represents over- and blue under-expression.
FIGURE 18.4 (Top): Error curves for the SRBCT data.
FIGURE 18.5. Regularized logistic regression paths for the leukemia data. The left panel is the lasso path, the right panel the elastic-net path with $\alpha = 0.8$. At the ends of the path (extreme left), there are 19 nonzero coefficients for the lasso, and 39 for the elastic net. The averaging effect of the elastic net results in more non-zero coefficients than the lasso, but with smaller magnitudes.
FIGURE 18.6. Training, test, and 10-fold cross validation curves for lasso logistic regression on the leukemia data. The left panel shows misclassification errors, the right panel shows deviance.
FIGURE 18.7. Protein mass spectrometry data: average profiles from normal and prostate cancer patients.
**Figure 18.8.** Fused lasso applied to CGH data. Each point represents the copy-number of a gene in a tumor sample, relative to that of a control (on the log base-2 scale).
FIGURE 18.9. Cross-validated ROC curves for protein example using the string kernel. The numbers next to each method in the legend give the area under the curve, an overall measure of accuracy. The SVM achieves better sensitivities than the other two, which achieve better specificities.
**FIGURE 18.10.** Abstracts example: top 20 scores from nearest shrunken centroids. Each score is the standardized difference in frequency for the word in the given class (BE, HT or JF) versus all classes. Thus a positive score (to the right of the vertical grey zero lines) indicates a higher frequency in that class; a negative score indicates a lower relative frequency.
FIGURE 18.11. Censored survival data. For illustration there are four patients. The first and third patients die before the study ends. The second patient is alive at the end of the study (365 days), while the fourth patient is lost to follow-up before the study ends. For example, this patient might have moved out of the country. The survival times for patients two and four are said to be “censored.”
FIGURE 18.12. Lymphoma data. The Kaplan–Meier estimate of the survival function for the 80 patients in the test set, along with one-standard-error curves. The curve estimates the probability of surviving past $t$ months. The ticks indicate censored observations.
FIGURE 18.13. Underlying conceptual model for supervised principal components. There are two cell types, and patients with the good cell type live longer on the average. Supervised principal components estimate the cell type, by averaging the expression of genes that reflect it.
FIGURE 18.14. Supervised principal components on the lymphoma data. The left panel shows a heatmap of a subset of the gene-expression training data. The rows are ordered by the magnitude of the univariate Cox-score, shown in the middle vertical column. The top 50 and bottom 50 genes are shown. The supervised principal component uses the top 27 genes (chosen by 10-fold CV). It is represented by the bar at the top of
FIGURE 18.15. Heatmap of the outcome (left column) and first 500 genes from a realization from model (16.37). The genes are in the columns, and the samples are in the rows.
FIGURE 18.16. Root mean squared test error (± one standard error), for supervised principal components and thresholded PLS on 100 realizations from model (16.37). All methods use one component, and the errors are relative to the noise standard deviation (the Bayes error is 1.0). For both methods, different values for the filtering threshold were tried and the number of features retained is shown on the horizontal axis. The extreme right points correspond to regular principal components and partial least squares, using all the genes.
FIGURE 18.17. Test errors for the lasso, supervised principal components, and pre-conditioned lasso, for one realization from model (16.37). Each model is indexed by the number of non-zero features. The supervised principal component path is truncated at 250 features. The lasso self-truncates at 100, the sample size (see Section 16.4). In this case, the pre-conditioned lasso achieves the lowest error with about 25 features.
FIGURE 18.18. Radiation sensitivity microarray example. A histogram of the 12,625 $t$-statistics comparing the radiation-sensitive versus insensitive groups. Overlaid in blue is the histogram of the $t$-statistics from 1000 permutations of the sample labels.
FIGURE 18.19. Microarray example continued. Shown is a plot of the ordered p-values $p(j)$ and the line $0.15 \cdot (j/12, 625)$, for the Benjamini–Hochberg method. The largest $j$ for which the p-value $p(j)$ falls below the line, gives the BH threshold. Here this occurs at $j = 11$, indicated by the vertical line. Thus the BH method calls significant the 11 genes (in red) with smallest p-values.