Today
1) Finish SVMs
2) Exam review

SVMs:
Reminder of last time
- Margin

Assume data is **linearly separable**
- Dual problem
  (solve for $\alpha$'s rather than $w$'s)
- Kernels $K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$
  Shortcut to projecting data into higher dimensional space

---

not linearly separable
Another way of dealing with overlap of data: Slack variables

Hard margin (linearly separable)
\[ \min_{\mathbf{w}} \frac{1}{2} \| \mathbf{w} \|^2 \]
\[ \text{s.t. } y_i (\mathbf{w}^T \mathbf{x}_i + b) \geq 1 + \xi_i \]

Soft margin
\[ \min_{\mathbf{w}} \frac{1}{2} \| \mathbf{w} \|^2 + C \sum_{i=1}^{N} \xi_i \]
\[ \text{s.t. } y_i (\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \xi_i + \xi_i \]
\[ \xi_i \geq 0 \quad \text{slack} \]

Dual \[ 0 \leq a_i < C \]
Practical issues

- How do you choose C? k?
  Cross validation or held-out data

- Number of support vectors → computational complexity

- So far we've only done 2 classes
  What about K classes? (K > 2)
  i vs. not i & K decisions → voting
  \((\binom{K}{2})\) i/j distinctions
  K-class extensions

- SVMs can also be used for regression (see HTF)

- Methods for getting confidences:

\[
p(y=+1 | \pi) = \frac{1}{1 + e^{-\sum_{i} a_i k(\pi_i, \pi)}}
\]
Exam Review

Exam covers everything covered in class 4 through model selection.

What I hope you've learned

tools: MAP rule, max likelihood est, EM algorithm, KNN, perceptron, classifier comparison

terminology: Bayes error rate, MAP, min risk, sufficient statistics, ROC, LDA, KNN, VC, bias/variance

practical considerations:

data a model complexity trade-offs
model complexity a computation trade-offs
bias/variance trade-offs for different model types

tuning parameters on held-out data
model assumptions
Basic strategy of classifier (or regression) design

0) What is the objective?
   classification vs. regression
   min error, weighted error, ...

1) Extract some features $X$
   Normalize features
   Optionally reduce features (PCA, LDA, filter selection)

2) Do we know the true probability distribution $p(x, y)$?
   YES $\Rightarrow$ get Bayes optimal result
   NO

   Assume model form
   (Gaussian, mixture, Gaussian mixture, exponent, multivariate, ...)
   Learn model parameter
   MLE (incl. EM), MMSE, MAP, Bayesian learning

   Assume classifier form
   (linear, quadratic, SVM, neural net)
   Learn function parameters
   Perception

   Use training data directly
   KNN
3) Improve the model via:
   model selection (Bayesian, CV)
   parameter tuning (CV)
   feature transformation, selection
   PCA, LOA
   \[\text{wrapper}\]

\[\text{Back to (2)}\]

4) Evaluation:
   error counting
   confidence intervals
   comparing classifiers
What are the sources of error in a classifier?

- overlap in feature space (measurement noise) = Bayes error
- non-representative training set
- overfitting
- bug
- wrong model assumptions
- random sample bad luck in training, finite data
- approximate solutions
Miscellaneous things

equivalence of different approaches:

\( x \sim N(\mu_i, \sigma^2 I) \)
linear classifier
min Euclidean distance to \( \mu_i \)
matched filter

\( x \sim N(\mu_i, \Sigma) \)
linear
min Mahalanobis dist

\( x \sim N(\mu_i, \Sigma_i) \)
quadric

\( p(x|wd) \) \( p(x| dl) \)

EER

\( d_{0,0} > d_{0,1} \)