ECE 5424: Introduction to Machine Learning

Topics:

- SVM
  - Lagrangian Duality
  - (Maybe) SVM dual & kernels

Readings: Barber 17.5

Stefan Lee
Virginia Tech
Recap of Last Time
Pittsburgh Welcomes Protest!

SAFER DATA MINING

FREE VARIABLES!

Baysians AGAINST

SUPPORT VECTOR MACHINES

Map Reduce, Map Reuse, Map Recycle

Those who oppress the poor and give to the rich will surely go.

Image Courtesy: Arthur Gretton
Generative vs. Discriminative

• Generative Approach (Naïve Bayes)
  – Estimate $p(x|y)$ and $p(y)$
  – Use Bayes Rule to predict $y$

• Discriminative Approach
  – Estimate $p(y|x)$ directly (Logistic Regression)
  – Learn “discriminant” function $f(x)$ (Support Vector Machine)
SVMs are Max-Margin Classifiers

Maximize this while getting examples correct.
Last Time

• Hard-Margin SVM Formulation

$$\min_{w,b} \quad w^T w$$

$$s.t. \quad (w^T x_j + b) y_j \geq 1, \forall j$$

• Soft-Margin SVM Formulation

$$\min_{w,b} \quad w^T w + C \sum_i \xi_i$$

$$s.t. \quad (w^T x_j + b) y_j \geq 1 - \xi_j, \forall j$$

$$\xi_j \geq 0, \quad \forall j$$
Last Time

• Can rewrite as a hinge loss

\[
\min_{w,b} \frac{1}{2} w^T w + C \sum_i h(y_i (w^T x_i + b))
\]
Today

• I want to show you how useful SVMs really are by explaining the Kernel trick but to do that….

• … we need to talk about Lagrangian Duality
Constrained Optimization

• How do I solve these sorts of problems:

\[
\begin{align*}
\min_{w} & \quad f(w) \\
\text{s.t.} & \quad h_1(w) = 0 \\
& \quad \vdots \\
& \quad h_k(w) = 0 \\
& \quad g_1(w) \leq 0 \\
& \quad \vdots \\
& \quad g_m(w) \leq 0
\end{align*}
\]
Introducing the Lagrangian

• Lets assume we have this problem

\[
\begin{align*}
\min_{w} & \quad f(w) \\
\text{s.t.} & \quad h(w) = 0 \\
& \quad g(w) \leq 0
\end{align*}
\]

Define the Lagrangian function:

\[
L(w, \beta, \alpha) = f(w) + \beta h(w) + \alpha g(w)
\]

\[
s.t. \quad \alpha \geq 0
\]

Where \(\beta\) and \(\alpha\) are called Lagrange Multipliers
(or dual variables)
Gradient of Lagrangian

\[ L(w, \beta, \alpha) = f(w) + \beta h(w) + \alpha g(w) \]

\[ \frac{\delta L}{\delta \alpha} = g(w) = 0 \]

\[ \frac{\delta L}{\delta \beta} = h(w) = 0 \]

\[ \frac{\delta L}{\delta w} = \frac{\delta f(w)}{\delta w} + \beta \frac{\delta h(w)}{\delta w} + \alpha \frac{\delta g(w)}{\delta w} = 0 \]

This will find **critical points** in the constrained function.
Building Intuition

\[ L(w, \beta, \alpha) = f(w) + \beta h(w) \]

\[ \frac{\delta L}{\delta w} = \frac{\delta f(w)}{\delta w} + \beta \frac{\delta h(w)}{\delta w} = 0 \]

\[ \frac{\delta f(w)}{\delta w} = -\beta \frac{\delta h(w)}{\delta w} \]
Geometric Intuition
Geometric Intuition

\[ f(x,y) = d_1 \]

\[ h'(x,y) = c \]

\[ f(x,y) = d_2 \]
A simple example

• Minimize $f(x,y) = x + y$ s.t. $x^2 + y^2 = 1$

• $L(x,y, \beta) = x + y + \beta(x^2 + y^2 - 1)$

• $\frac{\delta L}{\delta x} = 1 + 2\beta x = 0, \frac{\delta L}{\delta x} = 1 + 2\beta y = 0$

• $\frac{\delta L}{\delta B} = x^2 + y^2 - 1 = 0$

• $x = y = \frac{1}{2\beta} \rightarrow \frac{2}{4\beta^2} = 1 \rightarrow \beta = \pm\sqrt{2}$

• Makes $x = y = \pm\frac{\sqrt{2}}{2}$
Why go through this trouble?

- Often solutions derived from the Lagrangian form are easier to solve

- Sometimes they offer some useful intuition about the problem

- It builds character.
Lagrangian Duality

• More formally on overhead
  – Starring a game-theoretic interpretation, the duality gap, and KKT conditions.