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

Image Courtesy: Arthur Gretton

Image Courtesy: Arthur Gretton

Image Courtesy: Arthur Gretton

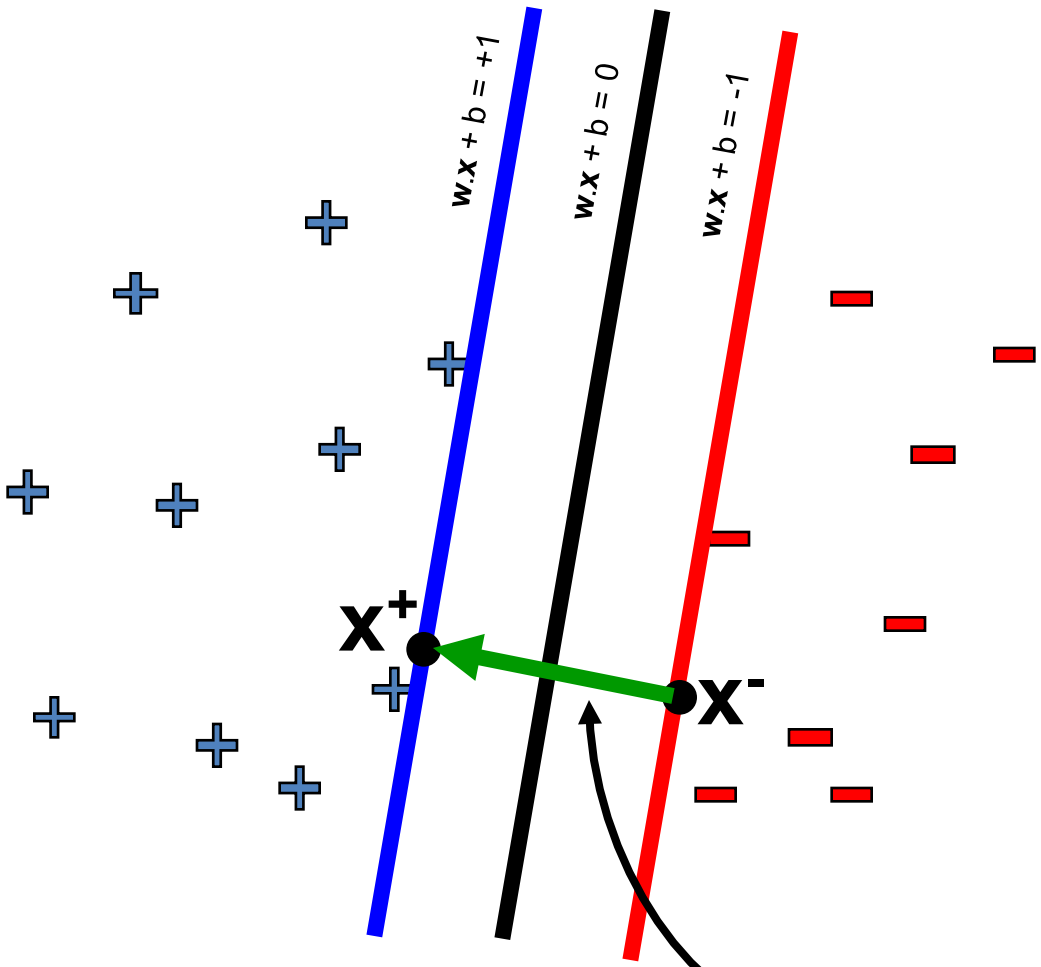
Image Courtesy: Arthur Gretton

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

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

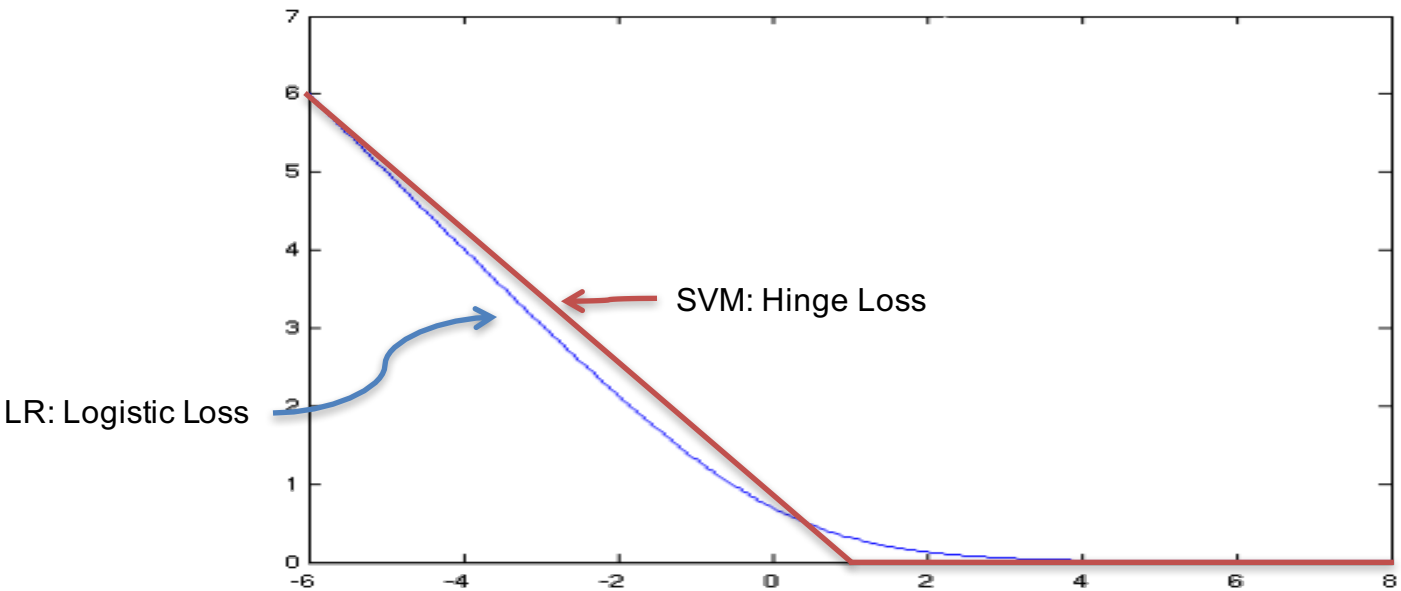
- Soft-Margin SVM Formulation

$$\min_{w,b} \quad w^T w + C \sum_i \xi_i$$
$$s.t. \quad \left(w^T x_j + b\right) y_j \geq 1-, \xi_j \quad \forall j$$
$$\xi_j \geq 0, \qquad \forall j$$

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:

$$
\begin{aligned}
&\min_{w} f(w) \\
&s.t. \quad h_1(w) = 0 \\
&\qquad\qquad \vdots \\
&\qquad\quad h_k(w) = 0 \\
&\qquad\quad g_1(w) \leq 0 \\
&\qquad\qquad \vdots \\
&\qquad\quad g_m(w) \leq 0
\end{aligned}
$$

Introducing the Lagrangian

- Lets assume we have this problem

$$\min_{w} f(w)$$
$$s.t. \quad h(w) = 0$$
$$g(w) \leq 0$$

Define the Lagrangian function:

$$L(w, \beta, \alpha) = f(w) + \beta h(w) + \alpha g(w)$$
$$s.t. \quad \alpha \geq 0$$

Where β and α 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

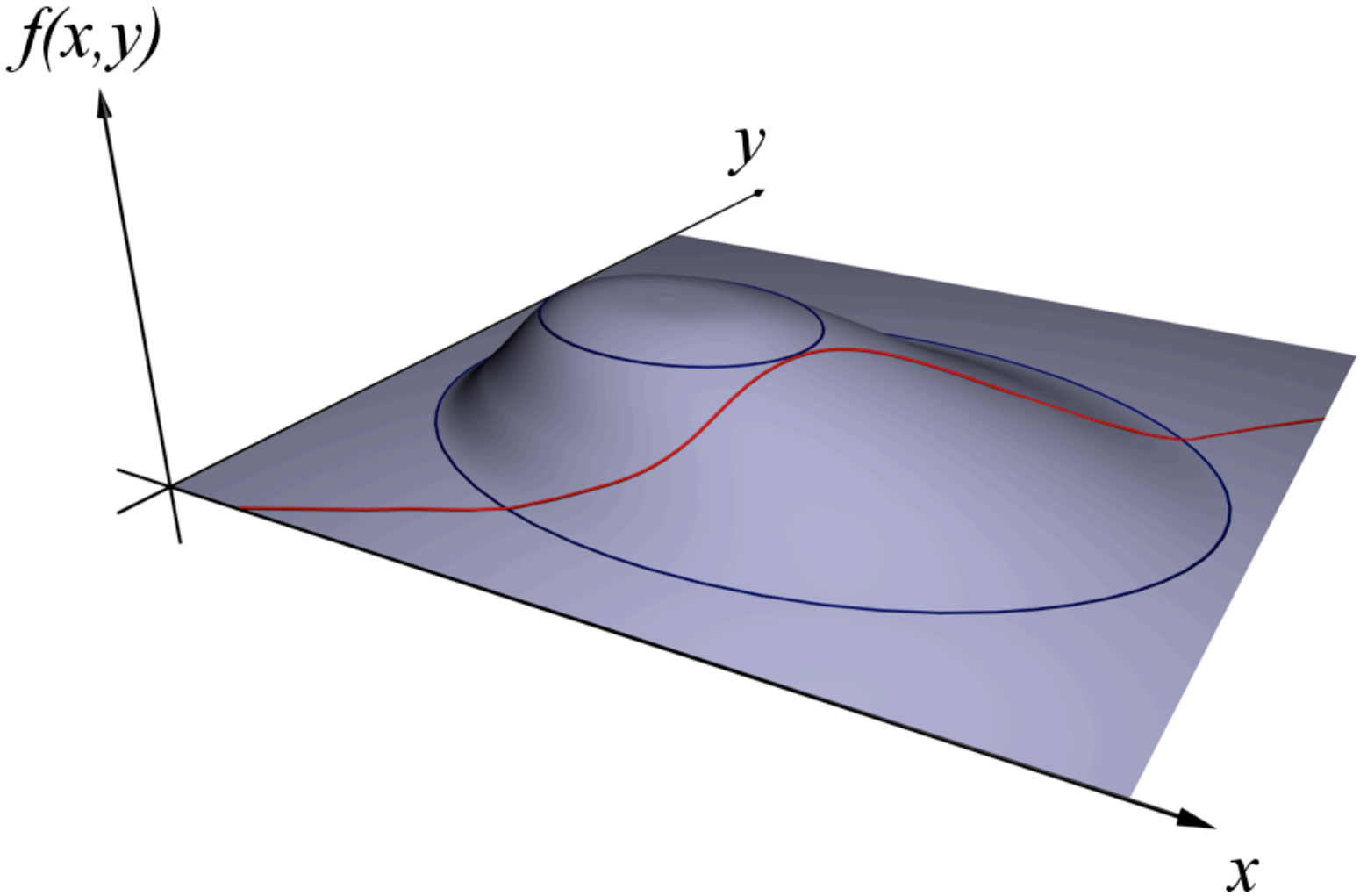


Image Credit: Wikipedia

Geometric Intuition

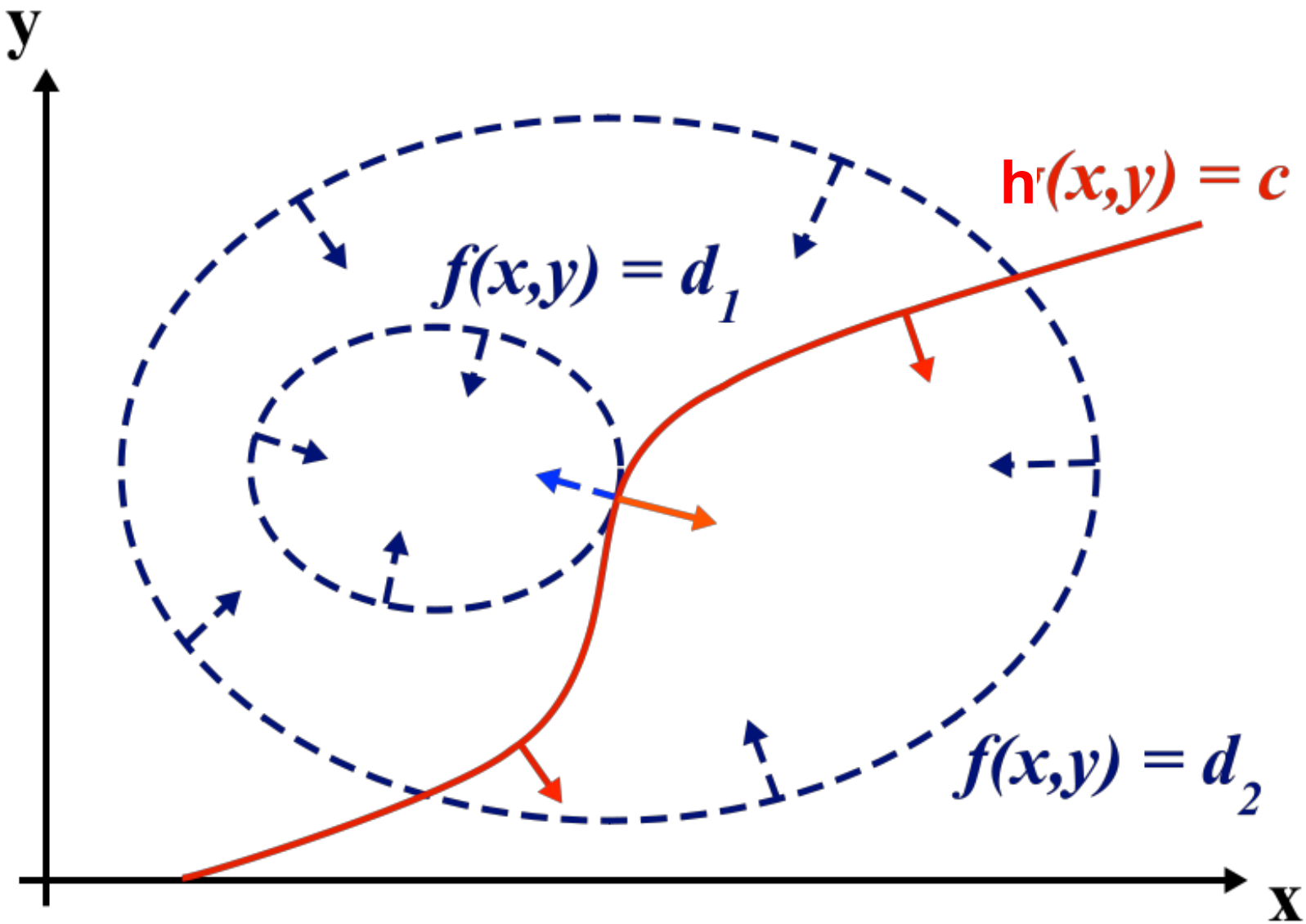
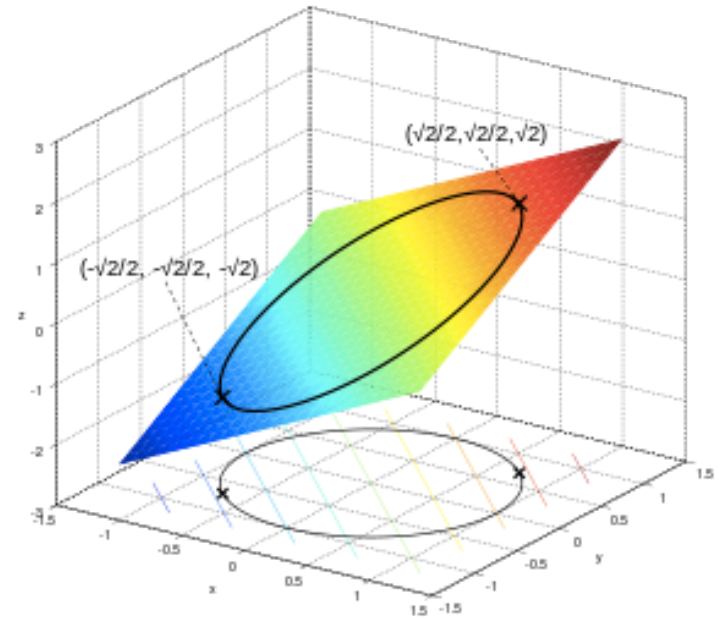


Image Credit: Wikipedia

A simple example

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



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

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

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

- 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.