Outline for March 27

- Finish Lagrangian for SVMs
- Kernels
- Soft-margin SVMs
Admin

* Last call for partners for Lab 6
* Final project presentations
  - Thurs May 16
    9-12 pm
  - (tentatively)
    Mon May 13
    4-6 pm
* Lab 3 back Friday!
* Lab 6: SVMs
* Lab 7: Deep Learning

Handout 11, Q2

\[ g(x, y) = 0 \]

Both maximize \( f(x, y) \)

Both 1.5 weeks
Outline for March 27

• Finish Lagrangian for SVMs
  • Kernels
  • Soft-margin SVMs
SVM (Goal)

\[
\begin{align*}
\min_{\hat{\omega}, b} & \quad \frac{1}{2} \|\hat{\omega}\|^2 \\
\text{s.t.} & \quad -y_i (\hat{\omega} \cdot \vec{x}_i + b) + 1 \leq 0 \\
& \quad i = 1, 2, \ldots, n \\
& \quad \hat{f} = 1 \\
& \quad \text{functional margin}
\end{align*}
\]

Lagrangian

\[
L(\omega, b, \alpha) = \frac{1}{2} \|\omega\|^2 - \sum_{i=1}^{n} \alpha_i [y_i (\omega \cdot \vec{x}_i + b) - 1]
\]

\[
\nabla L = 0 \quad \text{primal vs dual}
\]

\[
P^* = \min_{\omega, b} \left[ \max_{\alpha, \alpha \geq 0} L(\omega, b, \alpha) \right]
\]

\[
\Rightarrow \alpha^*, \alpha^* \Rightarrow \max_{\omega, b} \left[ \min_{\alpha, \alpha \geq 0} L(\omega, b, \alpha) \right] = d^*
\]

Sometimes

\[
P^* = d^*
\]
Solving the Dual

\( \nabla h(\omega, b, \alpha) = \omega - \sum_{i=1}^{n} \alpha_i y_i \hat{x}_i = 0 \)

\( \Rightarrow \quad \omega = \sum_{i=1}^{n} \alpha_i y_i \hat{x}_i \)

Like the perceptron updates!

\( \frac{\partial h(\omega, b, \alpha)}{\partial b} = -\sum_{i=1}^{n} \alpha_i y_i = 0 \)

\( \Rightarrow \quad \sum_{i=1}^{n} \alpha_i y_i = 0 \)

\( \alpha_i = 0 \) if \( \hat{x}_i \) not on margin

\( \alpha_i > 0 \) if \( \hat{x}_i \) is on margin

\( \sum_{i=1}^{n} \alpha_i = \sum \alpha_i \)

\( i : y_i = 1 \)

\( i : y_i = -1 \)

hyperplane
\[ h(w, b, \alpha) = \sum_{i=1}^{n} a_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j y_i y_j \] \[ x_i \cdot x_j \] \[ = W(\alpha) \]

Notice: no more w's + b's!

New dual optimization problem:

\[ \max_{\alpha} W(\alpha) \]
\[ \text{st. } \alpha_i \geq 0 \]
\[ \text{and } \sum_{i=1}^{n} \alpha_i y_i = 0 \]

Intuition: mean of similarities.
\[ h(\vec{w}, b, \vec{x}) = \sum_{i=1}^{n} a_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} y_i y_j a_i a_j \| \vec{x}_i \cdot \vec{x}_j \| = W(\vec{a}) \]

Notice: no more \( \vec{w} \)'s and \( b \)'s!

New dual optimization problem:

\[
\max_{\vec{a}} \quad W(\vec{a})
\]

\[
\text{s.t.} \quad a_i \geq 0
\]

and

\[
\sum_{i=1}^{n} a_i y_i = 0
\]

Intuition: measure of similarity

dot product is an example of an inner product
Outline for March 27

• Finish Lagrangian for SVMs
• Kernels
• Soft-margin SVMs
• How to solve the optimization problem
Kernel Idea

• By solving the dual form of the problem, we have seen how all computations can be done in terms of inner products between examples

• One example of an inner product is the dot product, which is the linear version of SVMs

• But there are many others!

• Intuition: if points are close together, their kernel function will have a large value (measure of similarity)
feature mapping

\[ \phi(\vec{x}) = \begin{bmatrix} x_1 \\ x_2 \\ x_1^2 + x_2^2 \end{bmatrix} \quad \vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \]

Kernel function

\[ K(\vec{x}, \vec{z}) = \phi(\vec{x}) \cdot \phi(\vec{z}) = \begin{bmatrix} x_1 \\ x_2 \\ x_1^2 + x_2^2 \end{bmatrix} \cdot \begin{bmatrix} z_1 \\ z_2 \\ z_1^2 + z_2^2 \end{bmatrix} = \vec{x} \cdot \vec{z} + \|\vec{x}\|^2 \|\vec{z}\|^2 \]
Kernel Trick example

Feature mapping: \( \varphi(x) = (x_1, x_2, x_1^2 + x_2^2) \)

Kernel function: \( K(x, z) = x \cdot z + \|x\|^2 \|z\|^2 \)

Original feature space

Mapping after applying kernel (can now find a hyperplane)
Gaussian Kernel

- Gaussian kernel is near 0 when points are far apart and near 1 when they are similar
- Also called Radial Basis Function (RBF) kernel

\[ K(\vec{x}, \vec{z}) = \exp \left( -\frac{||\vec{x} - \vec{z}||^2}{2\sigma^2} \right) \]
Gaussian Kernel

• Gaussian kernel is near 0 when points are far apart and near 1 when they are similar
• Also called Radial Basis Function (RBF) kernel

\[ K(\vec{x}, \vec{z}) = \exp \left( -\frac{||\vec{x} - \vec{z}||^2}{2\sigma^2} \right) \]

Often re-parametrized by gamma (different gamma!)

\[ \gamma = \frac{1}{2\sigma^2} \]

We will tune gamma as part of Lab 6
\[ h(\tilde{x}) = \text{sign} \left( \sum_{i=1}^{n} \alpha_i^* y_i \tilde{x} \cdot \tilde{x}_i + b \right) \]

Only need to consider the support vectors

Your Kernel function of choice

Kernel can be used for prediction as well!
Outline for March 27

• Finish Lagrangian for SVMs
• Kernels
• Soft-margin SVMs
Soft-margin SVMs (non-separable case)

• Idea: we will use regularization to add a cost for each point being incorrectly classified by the hyperplane

• Hopefully many costs will be 0, but we can accommodate a few outliers
Soft-margin SVMs (non-separable case)

- New optimization problem with regularization
- We will tune the $C$ parameter as part of Lab 6

\[
\begin{align*}
\min_{\xi, \bar{w}, b} & \quad \frac{1}{2} \| \bar{w} \|^2 + C \sum_{i=1}^{n} \xi_i \\
\text{s.t.} & \quad y_i (\bar{w} \cdot \bar{x}_i + b) \geq 1 - \xi_i, \quad i = 1, \ldots, n \\
\text{and} & \quad \xi_i \geq 0, \quad i = 1, \ldots, n
\end{align*}
\]