EECS 545: Machine Learning

Lecture 10. Kernel methods: Kernelizing Support Vector Machines

Honglak Lee

02/14/2020

Last update: 02/17/2020

Overview

- Support Vector Machine (SVM)
- Dual optimization
  - General recipe for constrained optimization
  - hard-margin SVM
  - soft-margin SVM

Maximum Margin Classifier

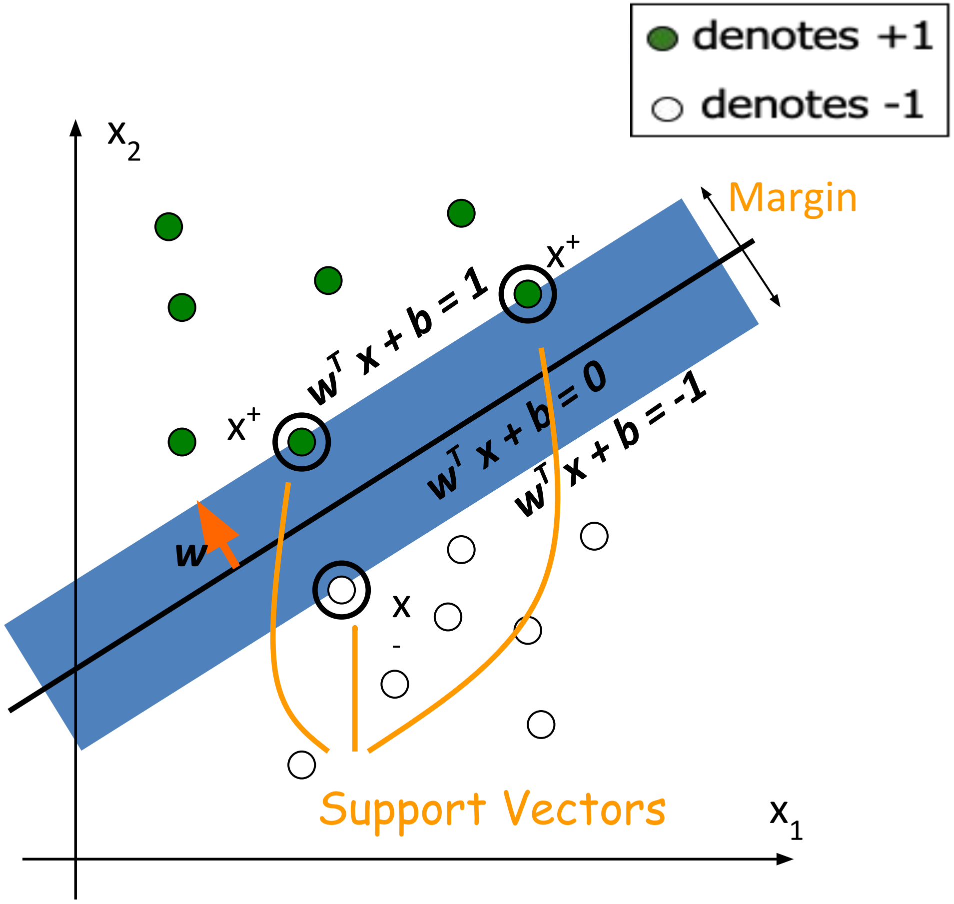
- Optimization problem:

$$\min_{\mathbf{w},b} \frac{1}{2}\|\mathbf{w}\|^2$$

subject to

For $y^{(n)} = 1,\ \mathbf{w}^T\phi(\mathbf{x}^{(n)}) + \mathrm{b} \geq 1$

For $y^{(n)} = -1, \mathbf{w}^T\phi(\mathbf{x}^{(n)}) + b \leq -1$



Dual optimization

- So far, we have considered primal optimization which requires a direct access to the feature vectors $\phi(\mathbf{x}^{(n)})$
- It is also possible to "kernelize" SVM
  - This formulation is called "Dual" formulation.
  - In this case, you can use any kernel function (such as polynomial, RBF, etc.)

With dual variables *a*$^{(n)}$, we have the following relations (without proofs)

$$\mathbf{w} = \sum_{n=1}^{N} a^{(n)} y^{(n)} \phi\left(\mathbf{x}^{(n)}\right)$$

$$h\left(\mathbf{x}\right) = \mathbf{w}^T \phi\left(\mathbf{x}\right) + b = \sum_{n=1}^{N} a^{(n)} y^{(n)} k\left(\mathbf{x}, \mathbf{x}^{(n)}\right) + b$$

Kernelizing SVM: back to hard-margin case

- Optimization problem:

$$\min_{\mathbf{w},b} \frac{1}{2}\|\mathbf{w}\|^2$$

subject to $\quad y^{(n)}\left(\mathbf{w}^T\phi\left(\mathbf{x}^{(n)}\right)+b\right)\geq 1, n=1,\cdots,N.$

- This is a constrained optimization problem.
  - We solve this using Lagrange multipliers (convex optimization)
  - Solving dual optimization problem naturally leads to kernalization

Solving Constrained Optimization: General Overview and Recipe

Constrained Optimization

- General **constrained problem** has the form:

$$\begin{aligned} \min_{\mathbf{x}} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & g_i(\mathbf{x}) \le 0,\ i = 1, ..., m \\ & h_i(\mathbf{x}) = 0,\ i = 1, ..., p \end{aligned}$$

- If **x** satisfies all the constraints, **x** is called <u>feasible</u>.

Lagrangian Formulation

- The **Lagrangian function** is

$$\mathcal{L}(\mathbf{x}, \lambda, \nu) = f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}) + \sum_{i=1}^{p} \nu_i h_i(\mathbf{x})$$

  - Here, $\lambda = [\lambda_1, \ldots, \lambda_m]$ $(\lambda_i \geq 0, \forall i)$ and $\nu = [\nu_1, \ldots, \nu_p]$ are called Lagrange multipliers (or dual variables)

- This leads to **primal optimization problem** (see next slide):

$$\min_{\mathbf{x}} \max_{\nu, \lambda : \lambda_i \geq 0, \forall i} \mathcal{L}(\mathbf{x}, \lambda, \nu)$$

- Difficult to solve directly!

Primal and Feasibility

- Primal optimization problem:

$$p^* = \min_{\mathbf{x}} \max_{\nu,\lambda:\lambda_i \geq 0, \forall i} \mathcal{L}(\mathbf{x}, \lambda, \nu)$$

  - where

$$\mathcal{L}(\mathbf{x}, \lambda, \nu) = f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}) + \sum_{i=1}^{p} \nu_i h_i(\mathbf{x})$$

- Notice that:

$$\mathcal{L}_p(\mathbf{x}) = \max_{\nu,\lambda:\lambda_i \geq 0, \forall i} \mathcal{L}(\mathbf{x}, \lambda, \nu) = \begin{cases} f(\mathbf{x}) & \text{if x is feasible} \\ \infty & \text{otherwise} \end{cases}$$

Lagrange Dual

- Dual optimization problem:

$$\max_{\nu,\lambda:\lambda_i \geq 0, \forall i} \min_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \lambda, \nu)$$

- We can also write as:

$$\begin{aligned} \max_{\lambda,\nu} \min_{\mathbf{x}} \quad & \mathcal{L}(\mathbf{x}, \lambda, \nu) \\ \text{subject to} \quad & \lambda_i \geq 0, \forall i \end{aligned}$$

Weak Duality

- Claim:
$$\begin{aligned} d^* &= \max_{\lambda,\nu:\lambda_i \geq 0} \min_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \lambda, \nu) \\ &\leq \min_{\mathbf{x}} \max_{\lambda,\nu:\lambda_i \geq 0} \mathcal{L}(\mathbf{x}, \lambda, \nu) \\ &= p^* \end{aligned}$$

- Difference between *p** and *d** is called <u>duality gap</u>.

Weak Duality

- Proof:

Let $\tilde{\mathbf{x}}$ be feasible. Then for any λ, ν with $\lambda_i \geq 0$,

$$\mathcal{L}(\tilde{\mathbf{x}}, \lambda, \nu) = f(\tilde{\mathbf{x}}) + \sum_i \lambda_i g_i(\tilde{\mathbf{x}}) + \sum_i \nu_i h_i(\tilde{\mathbf{x}}) \leq f(\tilde{\mathbf{x}})$$

Thus, $\tilde{\mathcal{L}}(\lambda, \nu) = \min_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \lambda, \nu) \leq \mathcal{L}(\tilde{\mathbf{x}}, \lambda, \nu) \leq f(\tilde{\mathbf{x}})$.
for any λ, ν with $\lambda_i \geq 0$, any feasible $\tilde{\mathbf{x}}$

Then,

$$d^* = \max_{\lambda, \nu : \lambda_i \geq 0} \tilde{\mathcal{L}}(\lambda, \nu) \leq f(\tilde{\mathbf{x}}) \text{ for any feasible } \tilde{\mathbf{x}}$$

Finally,

$$d^* = \max_{\lambda, \nu : \lambda_i \geq 0} \tilde{\mathcal{L}}(\lambda, \nu) \leq \min_{\tilde{\mathbf{x}}:\text{feasible}} f(\tilde{\mathbf{x}}) = p^*$$

Strong Duality

- If $p^* = d^*$, we say <u>strong duality</u> holds.

- What are the conditions for strong duality?
  - does not hold in general
  - holds for convex problems (under mild conditions)
  - conditions that guarantee strong duality in convex problems are called constraint qualification.

- Two well-known conditions
  - Slater’s constraint qualification
  - Karush-Kuhn-Tucker (KKT) condition

Conditions for strong duality: Slater's constraint qualification

- Strong duality holds for a convex problem

$$
\begin{aligned}
\min_{\mathbf{x}} \quad & f(\mathbf{x}) \\
\text{subject to} \quad & g_i(\mathbf{x}) \leq 0,\ i = 1, ..., m \\
& h_i(\mathbf{x}) = 0,\ i = 1, ..., p
\end{aligned}
$$

(where *f, g*$_i$ are convex, *and h*$_i$ are affine)

  – If it is <u>strictly</u> feasible, i.e.,

$$
\begin{aligned}
\exists x : \quad & g_i(\mathbf{x}) < 0,\ \forall i = 1, ..., m \\
& h_i(\mathbf{x}) = 0,\ \forall i = 1, ..., p
\end{aligned}
$$

Slater's condition is a sufficient condition for strong duality to hold for a convex problem

Karush-Kuhn-Tucker (KKT) condition

$$\nabla_{\mathbf{x}} f(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i \nabla_{\mathbf{x}} g_i(\mathbf{x}^*) + \sum_{i=1}^{p} \nu_i \nabla_{\mathbf{x}} h_i(\mathbf{x}^*) = 0 \quad (1)$$

$$h_i(\mathbf{x}^*) = 0,\ i = 1, \ldots, p \quad (2)$$

$$g_i(\mathbf{x}^*) \leq 0,\ i = 1, \ldots, m \quad (3)$$

$$\lambda_i^* \geq 0,\ i = 1, \ldots, m \quad (4)$$

$$\lambda_i^* g_i(\mathbf{x}^*) = 0,\ i = 1, \ldots, m \quad (5)$$

– The last condition is called complementary slackness.

Conditions for strong duality: KKT Conditions

- Assume *f, g_i, h_i* are differentiable

- If the original problem is **convex** (where *f, g_i* are convex, *and h_i* are affine) and $\mathbf{x}^*, \lambda^*, \nu^*$ satisfy the KKT conditions, then
  - $\mathbf{x}^*$ is primal optimal
  - (λ^*, ν^*) is dual optimal, and
  - the <u>duality gap is zero</u> (i.e., strong duality holds)

Proof for sufficiency

- From (2) and (3), $\mathbf{x}^*$ is primal feasible.
- From (4), (λ^*, ν^*) is dual feasible.
- $\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$ is a convex differentiable function. Thus, from (1), $\mathbf{x}^*$ is a minimizer of $\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$.
- Then,

$$
\begin{aligned}
d_0 \triangleq \tilde{\mathcal{L}}(\lambda^*, \nu^*) &= \min_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \lambda^*, \nu^*) \\
&= \mathcal{L}(\mathbf{x}^*, \lambda^*, \nu^*) \\
&= f(\mathbf{x}^*) + \sum_i \lambda_i g_i(\mathbf{x}^*) + \sum_i \nu_i h_i(\mathbf{x}^*) \\
&= f(\mathbf{x}^*)
\end{aligned}
$$

- Then,

$$d_0 \triangleq \widetilde{\mathcal{L}}(\lambda^*, \nu^*) \leq \underbrace{\max_{\lambda, \nu : \lambda_i \geq 0} \widetilde{\mathcal{L}}(\lambda, \nu) \leq \min_{\mathbf{x}\text{: feasible}} f(\mathbf{x})}_{\text{same proof as in weak duality}} \leq f(\mathbf{x}^*) = d_0$$

- Then,

$$\boxed{\max_{\lambda, \nu : \lambda_i \geq 0} \widetilde{\mathcal{L}}(\lambda, \nu) = \min_{\mathbf{x}\text{: feasible}} f(\mathbf{x})} = d_0$$

which proves that the strong duality holds (i.e., duality gap is zero).

KKT conditions: Conclusion

- If a constrained optimization if differentiable and has convex objective function and constraint sets, then the KKT conditions are **(necessary and) sufficient conditions** for **strong duality** (zero duality gap).

- Thus, the KKT conditions can be used to solve such problems.

Recap: General Recipe

- Given an original optimization

$$\begin{aligned} \min_{\mathbf{x}} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & g_i(\mathbf{x}) \leq 0,\ i = 1, ..., m \\ & h_i(\mathbf{x}) = 0,\ i = 1, ..., p \end{aligned}$$

- Solve dual optimization with Lagrangian function:

$$\begin{aligned} \max_{\lambda,\nu} \min_{\mathbf{x}} \quad & \mathcal{L}(\mathbf{x}, \lambda, \nu) = f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}) + \sum_{i=1}^{p} \nu_i h_i(\mathbf{x}) \\ \text{subject to} \quad & \lambda_i \geq 0,\ \forall i \end{aligned}$$

- Alternatively, solve the dual optimization with Lagrange dual:

$$\begin{aligned} \max_{\lambda,\nu} \quad & \tilde{\mathcal{L}}(\lambda, \nu) \qquad \text{where} \ \ \tilde{\mathcal{L}}(\lambda, \nu) = \min_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \lambda, \nu) \\ \text{subject to} \quad & \lambda_i \geq 0,\ \forall i \end{aligned}$$

Recap: KKT Optimality condition

- Karush-Kuhn-Tucker (KKT) condition:

$$\nabla_{\mathbf{x}} f(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i \nabla_{\mathbf{x}} g_i(\mathbf{x}^*) + \sum_{i=1}^{p} \nu_i \nabla_{\mathbf{x}} h_i(\mathbf{x}^*) = 0$$

$$h_i(\mathbf{x}^*) = 0,\ i = 1, ..., p$$

$$g_i(\mathbf{x}^*) \leq 0,\ i = 1, ..., m$$

$$\lambda_i^* \geq 0,\ i = 1, ..., m$$

$$\lambda_i^* g_i(\mathbf{x}^*) = 0,\ i = 1, ..., m$$

  – The last condition is called complementary slackness, and guarantees the strong duality for convex optimization

Applying Constrained Optimization Techniques for solving SVM

Kernelizing SVM: back to hard-margin case

- Optimization problem:

$$\min_{\mathbf{w},b} \frac{1}{2}\|\mathbf{w}\|^2$$

subject t $y^{(n)}\left(\mathbf{w}^T\phi\left(\mathbf{x}^{(n)}\right)+b\right)\geq 1, n=1,\cdots,N.$

- This is a constrained optimization problem.
  - We solve this using Lagrange multipliers (convex optimization)

Back to hard-margin SVM

- Use Lagrange multipliers to enforce constraints while optimizing

$$L(\mathbf{w}, b, a) = \frac{1}{2}\|\mathbf{w}\|^2 + \sum_{n=1}^{N} a^{(n)} \left\{1 - y^{(n)} \left(\mathbf{w}^T \phi\left(\mathbf{x}^{(n)}\right) + b\right)\right\}$$

- Here, $a^{(n)} \geq 0$ is the Lagrange multiplier (or dual variable) for each constraint

$$y^{(n)} \left(\mathbf{w}^T \phi\left(\mathbf{x}^{(n)}\right) + b\right) \geq 1 \quad n = 1, \ldots, N.$$

Lagrangian and Lagrange Dual

- Optimizing the <u>Lagrangian</u> :

$$\max_{\mathbf{a}} \min_{\mathbf{w},b} L(\mathbf{w}, b, a) = \frac{1}{2}\|\mathbf{w}\|^2 + \sum_{n=1}^{N} a^{(n)} \left\{1 - y^{(n)} \left(\mathbf{w}^T \phi\left(\mathbf{x}^{(n)}\right) + b\right)\right\}$$

subject to $a^{(n)} \geq 0, \forall n$

- We first minimize with respect to **w** and b, and get <u>Lagrange dual</u> $\max_{\mathbf{a}} \widetilde{L}(\mathbf{a})$

where $\widetilde{L}(\mathbf{a}) = \min_{\mathbf{w},b} L(\mathbf{w}, b, \mathbf{a})$

subject to $a^{(n)} \geq 0, \forall n$

Maximize the Margin

- Set the derivatives of $L(\mathbf{w}, b, \mathbf{a})$ to zero, to get

$$\mathbf{w} = \sum_{n=1}^{N} a^{(n)} y^{(n)} \phi\left(\mathbf{x}^{(n)}\right) \qquad 0 = \sum_{n=1}^{N} a^{(n)} y^{(n)}$$

- Substitute in, to eliminate w and *b*,

$$\max_{a} \widetilde{L}\left(\mathbf{a}\right) = \sum_{n=1}^{N} a^{(n)} - \frac{1}{2} \sum_{n=1}^{N} \sum_{m=1}^{N} a^{(n)} a^{(m)} y^{(n)} y^{(m)} \phi\left(\mathbf{x}^{(n)}\right)^{T} \phi\left(\mathbf{x}^{(m)}\right)$$

subject to $a^{(n)} \geq 0, \forall n$

Dual Representation (with kernel)

- Define a kernel $k\left(\mathbf{x}^{(n)}, \mathbf{x}^{(m)}\right) = \phi(\mathbf{x}^{(n)})^T \phi(\mathbf{x}^{(m)})$
- This gives, to maximize

$$\max_{a} \widetilde{L}(\mathbf{a}) = \sum_{n=1}^{N} a^{(n)} - \frac{1}{2} \sum_{n=1}^{N} \sum_{m=1}^{N} a^{(n)} a^{(m)} y^{(n)} y^{(m)} k\left(\mathbf{x}^{(n)}, \mathbf{x}^{(m)}\right)$$

subject to $a^{(n)} \geq 0, \forall n$

- Once we have **a**, we don't need **w**. Predict new values using:

$$h(\mathbf{x}) = \mathbf{w}^T \phi(\mathbf{x}) + b = \sum_{n=1}^{N} a^{(n)} y^{(n)} k\left(\mathbf{x}, \mathbf{x}^{(n)}\right) + b$$

Recovering b

- For any support vector $\mathbf{x}^{(n)}: y^{(n)} h\left(\mathbf{x}^{(n)}\right) = 1$
- Replacing with $h(\mathbf{x}) = \mathbf{w}^T \phi(\mathbf{x}) + b = \sum_{n=1}^{N} a^{(n)} y^{(n)} k\left(\mathbf{x}, \mathbf{x}^{(n)}\right) + b$

$$y^{(n)} \left(\sum_{m \in S} a^{(m)} y^{(m)} k\left(\mathbf{x}^{(n)}, \mathbf{x}^{(m)}\right) + b \right) = 1$$

(index) set of support vectors

- Multiply $y^{(n)}$, and sum over n:

$$b = \frac{1}{N_S} \sum_{n \in S} \left(y^{(n)} - \sum_{m \in S} a^{(m)} y^{(m)} k\left(\mathbf{x}^{(n)}, \mathbf{x}^{(m)}\right) \right)$$

Support Vectors

- The KKT conditions are:

$$\nabla_{\mathbf{w}} L(\mathbf{w}, b, \mathbf{a}) = 0$$
$$\nabla_b L(\mathbf{w}, b, \mathbf{a}) = 0$$
$$a^{(n)} \geq 0$$
$$1 - y^{(n)} h\left(\mathbf{x}^{(n)}\right) \leq 0$$
$$a^{(n)} \left\{1 - y^{(n)} h\left(\mathbf{x}^{(n)}\right)\right\} = 0$$

- The last condition means:
  - either or $a^{(n)} = 0$ or $y^{(n)} h(\mathbf{x}^{(n)}) = 1$.

- That is, only the support vectors matter!
  - To predict $h(\mathbf{x})$, sum only over support vectors

Soft SVM

- Maximize the margin, and also penalize for the slack variables

$$C \sum_{n=1}^{N} \xi_n + \frac{1}{2}||\mathbf{w}||^2$$

- The support vectors are now those with

$$y^{(n)} h\left(\mathbf{x}^{(n)}\right) = 1 - \xi^{(n)}$$

Formulation of soft-margin SVM

- Primal optimization
  - Optimization w.r.t. w and $\xi^{(n)}$'s:

$$\min_{\mathbf{w},b,\boldsymbol{\xi}} \quad C\sum_{n=1}^{N}\xi_n + \frac{1}{2}\|\mathbf{w}\|^2$$

subject to $$y^{(n)} h\left(\mathbf{x}^{(n)}\right) \geq 1 - \xi^{(n)}, \ \forall n$$
$$\xi^{(n)} \geq 0, \forall n$$

Dual formulation of soft-margin SVM

- Lagrangian

$$L(\mathbf{w}, b, \xi, a, \mu) = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{n=1}^{N}\xi^{(n)} + \sum_{n=1}^{N} a^{(n)}\left\{1 - y^{(n)}h\left(\mathbf{x}^{(n)}\right) - \xi^{(n)}\right\} + \sum_{n=1}^{N}\mu^{(n)}(-\xi^{(n)})$$

  – where $a^{(n)} \geq 0$, $\mu^{(n)} \geq 0$, $\xi^{(n)} \geq 0, \forall n$

- KKT conditions for the constraints

$$\left.\begin{aligned} 1 - y^{(n)}h\left(\mathbf{x}^{(n)}\right) - \xi^{(n)} &\leq 0 \\ -\xi^{(n)} &\leq 0 \end{aligned}\right\} \text{Primal variables satisfy the inequality constraints}$$

$$\left.\begin{aligned} a^{(n)} &\geq 0 \\ \mu^{(n)} &\geq 0 \end{aligned}\right\} \text{Dual variables (for above inequalities) are feasible}$$

$$\left.\begin{aligned} a^{(n)}\left(1 - y^{(n)}h\left(\mathbf{x}^{(n)}\right) - \xi^{(n)}\right) &= 0 \\ \mu^{(n)}\xi^{(n)} &= 0 \end{aligned}\right\} \text{Complementary slackness condition}$$

Dual formulation of soft-margin SVM

- Taking derivatives

$$\frac{\partial L}{\partial \mathbf{w}} = 0 \quad \Rightarrow \quad \mathbf{w} = \sum_{n=1}^{N} a^{(n)} y^{(n)} \phi\left(\mathbf{x}^{(n)}\right)$$

$$\frac{\partial L}{\partial b} = 0 \quad \Rightarrow \quad \sum_{n=1}^{N} a^{(n)} y^{(n)} = 0$$

$$\frac{\partial L}{\partial \xi_n} = 0 \quad \Rightarrow \quad a^{(n)} = C - \mu^{(n)}$$

Dual formulation of soft-margin SVM

- Dual optimization (via Lagrange dual)

$$\min_{\mathbf{a}} \widetilde{L}(\mathbf{a}) = \sum_{n=1}^{N} a^{(n)} - \frac{1}{2} \sum_{n=1}^{N} \sum_{m=1}^{N} a^{(n)} a^{(m)} y^{(n)} y^{(m)} k\left(\mathbf{x}^{(n)}, \mathbf{x}^{(m)}\right)$$

subject to $\quad 0 \leq a^{(n)} \leq C$

$$\sum_{n=1}^{N} a^{(n)} y^{(n)} = 0$$

- Solve quadratic problem (convex optimization)

SVM: practical issues

Support Vector Machine: Algorithm

1. Choose a kernel function

2. Choose a value for C

3. Solve the optimization problem (many software packages available) – primal or dual

4. Construct the discriminant function from the support vectors

Some Issues

- Linear kernels work fairly well, but can be suboptimal.

- Choice of (nonlinear) kernels
  – Gaussian or polynomial kernel is default
  – If the simple kernels are ineffective, more elaborate kernels are needed
  – Domain experts can give assistance in formulating appropriate similarity measures

- Choice of kernel parameters
  – E.g., Gaussian kernel: $K(\mathbf{x}, \mathbf{z}) = \exp\left(-\frac{\|\mathbf{x}-\mathbf{z}\|^2}{2\sigma^2}\right)$
    - σ is the distance between neighboring points whose labels will likely to affect the prediction of the query point.

  – In the absence of reliable criteria, applications rely on the use of a validation set or cross-validation to set such parameters.

Summary: Support Vector Machine

- Max Margin Classifier
  - Better generalization ability & less over-fitting
  - Solved by convex optimization techniques

- The Kernel Trick
  - Map data points to higher dimensional space in order to make them linearly separable.
  - Since only dot product is used, we do not need to represent the mapping explicitly.

Additional Resource

- Kernel Methods
  - http://www.kernel-machines.org/
- Convex Optimization
  - http://www.stanford.edu/~boyd/cvxbook/
  - http://www.stanford.edu/class/ee364a/
  - see Chapter 5 (and earlier chapters)

SVM Implementation

- LIBSVM
  - http://www.csie.ntu.edu.tw/~cjlin/libsvm/
  - One of the most popular generic SVM solver (supports nonlinear kernels)
- Liblinear
  - http://www.csie.ntu.edu.tw/~cjlin/liblinear/
  - One of the fastest linear SVM solver (linear kernel)
- SVMlight
  - http://www.cs.cornell.edu/people/tj/svm_light/
  - Structured outputs, various objective measure (e.g., F1, ROC area), Ranking, etc.

SVM demo code

- http://www.mathworks.com/matlabcentral/fileexchange/28302-svm-demo

- http://www.alivelearn.net/?p=912

Change logs

Updates: 2/17/2020

- For the SVM optimization definition (e.g., p.3 and later on), changed from “arg min” to “min”
- For the proof of strong duality from KKT conditions (p. 19) “min $f(\mathbf{x})$” is changed “min $f(\mathbf{x})$ subject to $\mathbf{x}$: feasible”