SVM review

And some feedback from the class

Is this a valid kernel?

Assume x = $[x_1, x_2]$

1) $k(\mathbf{x},\mathbf{x}') = x_1 + x_2 + x_1' + x_2'$

2) $k(\mathbf{x},\mathbf{x}') = x_1^2 + x_2^2 + x_1'^2 + x_2'^2$

3) $k(\mathbf{x},\mathbf{x}') = \exp((x_1 - x_1')^2 + (x_2 - x_2')^2)$

Is this a valid kernel?

Assume x = $[x_1, x_2]$

1) $k(\mathbf{x},\mathbf{x}') = x_1 + x_2 + x_1' + x_2'$

No – this is not positive semi-definite

2) $k(\mathbf{x},\mathbf{x}') = x_1^2 + x_2^2 + x_1'^2 + x_2'^2$

probablye – requires a more careful look

3) $k(\mathbf{x},\mathbf{x}') = \exp((x_1 - x_1')^2 + (x_2 - x_2')^2)$

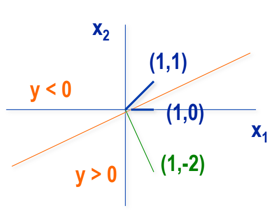
yes – this is the Gaussian kernel

Hyperplanes

- **Given the hyperplane defined by the line**
  - $y = x_1 - 2x_2$
  - $y = (1,-2)^T \mathbf{x} = \mathbf{w}^T \mathbf{x}$
- **Is this point correctly predicted?**
  - $y = 1$, $\mathbf{x} = (1,0)$?
  - $y = 1$, $\mathbf{x} = (1,1)$?
- **How should the weight be adjusted to reduce the loss function?**
  - .

Hyperplanes

- **Given the hyperplane defined by the line**
  - $y = x_1 - 2x_2$
  - $y = (1,-2)^T \mathbf{x}$



So (1,0) having y=1 is correct and (1,1) having y=1 is not correct

Hyperplanes

- ***True or False?* When solving for a hyperplane specified by** $\mathbf{w}^\top \mathbf{x} + b = 0$ **one can always set the margin to 1;**

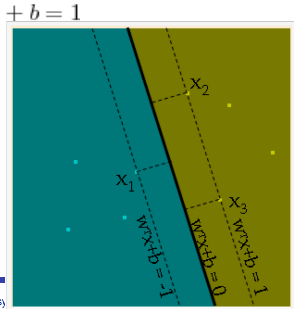
$$\mathbf{w}^\top \mathbf{x_1} + b = -1 \text{ and } \mathbf{w}^\top \mathbf{x_2} + b = 1$$

Hyperplanes

- *True or False?* **When solving for a hyperplane specified by $\mathbf{w}^\top \mathbf{x} + b = 0$ one can always set the margin to 1;**

$$\mathbf{w}^\top \mathbf{x}_1 + b = -1 \text{ and } \mathbf{w}^\top \mathbf{x}_2 + b = 1$$

true



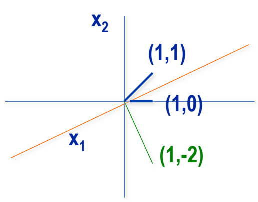
Projections

- **The projection of a point x onto a line w is**
  x^T w/|w|$_2$
- **The distance of a point x to a hyperplane defined by the line w is**
  x^T w/|w|$_2$

Why?

Hyperplanes

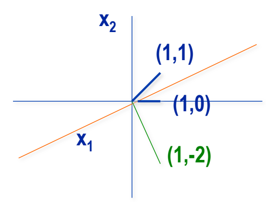
- **The projection of a point x on a plane defined by a line w is x^T w/|w|$_2$**
- **The distance of x from the hyperplane defined by (1, -2) is what**
  - for x = (-1,2)
  - for x = (1,0)
  - for x = (1,1)



Hyperplanes

- **The projection of a point x on a plane defined by a line w is x^T w/|w|$_2$**
- **The distance of x from the hyperplane defined by (1, -2) is what**
  - for x = (-1,2)
  - for x = (1,0)
  - for x = (1,1)

The same as before – you project onto the line



Hyperplanes

- *True or False?* **When solving for a hyperplane specified by $\mathbf{w}^\top \mathbf{x} + b = 0$ one can always set the margin to 1;**

$$\mathbf{w}^\top \mathbf{x}_1 + b = -1 \text{ and } \mathbf{w}^\top \mathbf{x}_2 + b = 1$$

- *True or False?* **This then implies that the margin is the distance of the support vectors from the separating hyperplane**

$$\frac{\mathbf{w}}{2||\mathbf{w}||_2}(\mathbf{x}_2 - \mathbf{x}_1) = \frac{1}{||\mathbf{w}||_2}$$

(non)separable SVMs

- ***True or False?*** In a real problem, you should check to see if the SVM is separable and then include slack variables if it is not separable.
- ***True or False?*** Adding slack variables is equivalent to requiring that all of the α_i are less than a constant in the dual

$$\mathbf{w}^\top \phi(\mathbf{x}) + b = \sum_i \alpha_i y_i k(\mathbf{x}_i, \mathbf{x}) + b$$

(non)separable SVMs

- *True or False?* In a real problem, you should check to see if the SVM is separable and then include slack variables if it is not separable.
  - False: you can just run the slack variable problem in either case
- *True or False?* Adding slack variables is equivalent to requiring that all of the α_i are less than a constant in the dual

$$\mathbf{w}^\top \phi(\mathbf{x}) + b = \sum_i \alpha_i y_i k(\mathbf{x}_i, \mathbf{x}) + b$$

true

Why do SVMs work well?

- **Why are SVMs fast?**
- **Why are SVMs often more accurate than logistic regression?**

Why do SVMs work well?

- **Why are SVMs fast?**
  - They work in the dual, with relatively few points
  - The kernel trick
- **Why are SVMs often more accurate than logistic regression?**
  - SVMs use kernels
    - but logistic regression can also do that
  - Logistic regression uses all the data points, assuming a probabilistic model, while SVMs ignore the points that are clearly correct.

Class feedback

- **What you liked**
  - **My enthusiasm, focus on intuition**
  - **The HW**
  - The subject
  - Clickers

What could be improved?

- **Make homework clearer**
  - fewer (maybe longer) HWs
  - More short questions to help clarify notation
  - latex is a pain
- **More math**
- **Less math**
- **More (applied) projects**
- **More theory**
- **Less work**

What could be improved?

- **More detailed explanation of clicker questions**
  - post answers
- **Some (extra) session where we do more of a "hands-down" approach to problems.**
- **Change exam questions to be more like homework**

Class feedback

- Speed
  - 36% **too fast**
  - 61% **good**
  - 3% **too slow**
- Average hours/week
  - 17
- Homeworks
  - All very close to "4 – probably keep"