PATTERN RECOGNITION
ASSIGNMENT 3.
FLD, LDF, SVM.
Due: Nov. 10, 2019.

Do the following problems. [1]

1. Problem 42, sect. 3.8, p. 153.
   Consider the criterion function $J(\mathbf{w})$ required for the Fisher linear discriminant function

   (a) (4 marks) Fill in the steps leading from Eqs. 96, 98, and 102 to Eq. 103, i.e., from

   $$J(\mathbf{w}) = \frac{|\tilde{m}_1 - \tilde{m}_2|}{\tilde{s}_1^2 + \tilde{s}_2^2} \tag{96}$$

   $$\mathbf{S}_W = \mathbf{S}_2 + \mathbf{S}_2 \tag{98}$$

   $$\mathbf{S}_B = (\mathbf{m}_1 - \mathbf{m}_2)(\mathbf{m}_1 - \mathbf{m}_2)^t \tag{102}$$

   to

   $$J(\mathbf{w}) = \frac{\mathbf{w}^t \mathbf{S}_B \mathbf{w}}{\mathbf{w}^t \mathbf{S}_W \mathbf{w}}. \tag{103}$$

   (b) (13 marks) Use matrix method to show that the solution to Eq. 103 is indeed given by Eq. 104, i.e.,

   $$\mathbf{S}_B \mathbf{w} = \lambda \mathbf{S}_W \mathbf{w} \tag{104}$$

   (c) (13 marks) At the extreme of $J(\mathbf{w})$, a small change in $\mathbf{w}$ must leave $J(\mathbf{w})$ unchanged. Consider a small perturbation away from the optimal, $\mathbf{w} + \Delta\mathbf{w}$, and derive the solution condition of Eq. 104.

2. (30 marks) Problem 19, sect. 5.8, p. 274.
   Given the conditions in Eqs. 28-30,

   $$\eta(k) \geq 0 \tag{28}$$

   $$\lim_{m\to\infty} \sum_{k=1}^{m} \eta(k) = \infty \tag{29}$$

   $$\lim_{m\to\infty} \frac{\sum_{k=1}^{m} \eta^2(k)}{\left(\sum_{k=1}^{m} \eta(k)\right)^2} = 0 \tag{30}$$

   show that $\mathbf{a}(k)$ in the variable increment descent rule indeed converges for $\mathbf{a}^t\mathbf{y}_i > b$ for all i.

3. Problem 32 (a), (b), sect. 5.11, p. 276.
   Consider a Support Vector Machine and the following training data from two categories:

   $$\omega_1 : (1,1)^t, (2,2)^t, (2,0)^t$$
   $$\omega_2 : (0,0)^t, (1,0)^t, (0,1)^t$$

   (a) (5 marks) Plot these six training points, and construct by inspection the weight vector for the optimal hyperplane, and the optimal margin.

   (b) (5 marks) What are the support vectors?

[1] see the textbook Duda, Hart, and Stork *Pattern Classification* 2nd edition, 2001

4. Problem 1, sect. 5.4, p. 278 (Computer Exercise).
   Consider basic gradient descent (Algorithm 1) and the Perceptron criterion (Eq. 16)

$$J_p(\mathbf{a}) = \sum_{y \in \mathcal{Y}} (-\mathbf{a}^{\mathbf{t}}\mathbf{y})$$

   applied to the data in the table

	ω_1		ω_2		ω_3		ω_4	
sample	x_1	x_2	x_1	x_2	x_1	x_2	x_1	x_2
1	0.1	1.1	7.1	4.2	-3.0	-2.9	-2.0	-8.4
2	6.8	7.1	-1.4	-4.3	0.5	8.7	-8.9	0.2
3	-3.5	-4.1	4.5	0.0	2.9	2.1	-4.2	-7.7
4	2.0	2.7	6.3	1.6	-0.1	5.2	-8.5	-3.2
5	4.1	2.8	4.2	1.9	-4.0	2.2	-6.7	-4.0
6	3.1	5.0	1.4	-3.2	-1.3	3.7	-0.5	-9.2
7	-0.8	-1.3	2.4	-4.0	-3.4	6.2	-5.3	-6.7
8	0.9	1.2	2.5	-6.1	-4.1	3.4	-8.7	-6.4
9	5.0	6.4	8.4	3.7	-5.1	1.6	-7.1	-9.7
10	3.9	4.0	4.1	-2.2	1.9	5.1	-8.0	-6.3

   (a) (20 marks) Apply both to the two-dimensional data in order to discriminate categories ω_1 and ω_2 and ω_1 and ω_4. For the gradient descent use $\eta(k) = 0.1$. Plot the criterion function as function of the iteration number.

   (b) (5 marks) Estimate the total number of mathematical operations in the two algorithms.

   (c) (5 marks) Plot the convergence time versus learning rate. What is the minimum learning rate that fails to lead to convergence?

5. (bonus 20 marks) Problem 32 (c), sect. 5.11, p. 276.

   (c) Construct the solution in the dual space by finding the Lagrange undetermined multipliers, α_i. Compare your result to that in part (a). are the support vectors? Provide full analytical solution to the problem.