Supplementary document: Learning hash functions using column generation*

Xi Li, Guosheng Li, Chunhua Shen, Anton van den Hengel, Anthony Dick
School of Computer Science, University of Adelaide

1. Overview

We provide more details in this supplementary document on (1) learning hash functions with the hinge loss; (2) learning hash functions with general convex loss. Those information is omitted in our paper [3] due to space limit.

More specifically, the optimization problems in (1) are regularized by two types of norms: l_1 norm and l_∞ norm. (2) We provide the details of deriving the optimization formulation using squared hinge loss and logistic loss regularized by the l_1 and l_∞ norms.

2. Learning hash functions with the hinge loss

First we consider the l_1 norm regularization.

2.1. Hashing with the l_1 norm regularization

Using the hinge loss, we can write the following large-margin optimization problem:

$$\begin{aligned} \min_{\boldsymbol{w},\boldsymbol{\xi}} \quad & \textstyle\sum_{i=1}^{|\mathcal{I}|} \xi_i + C\|\boldsymbol{w}\|_1 \\ \text{s.t.} \quad & \boldsymbol{w} \succcurlyeq \mathbf{0};\ d_{\mathcal{H}}(\mathbf{x}_i, \mathbf{x}_i^-) - d_{\mathcal{H}}(\mathbf{x}_i, \mathbf{x}_i^+) \geq 1 - \xi_i,\ \xi_i \geq 0,\ \forall i = 1 \cdots . \end{aligned} \tag{1}$$

Here $\|\cdot\|_1$ is the 1-norm; $\boldsymbol{w} = (w_1, w_2, \ldots, w_\ell)^\top$ is the weight vector; ξ_i is a slack variable; and C is a parameter controlling the trade-off between training error and model capacity.

For notational simplicity, we define

$$a_m^{[i]} = |h_m(\mathbf{x}_i) - h_m(\mathbf{x}_i^-)| - |h_m(\mathbf{x}_i) - h_m(\mathbf{x}_i^+)|.$$

Then $d_{\mathcal{H}}(\mathbf{x}_i, \mathbf{x}_i^-) - d_{\mathcal{H}}(\mathbf{x}_i, \mathbf{x}_i^+) = \boldsymbol{w}^\top \mathbf{a}_i$ with $\mathbf{a}_i = (a_1^{[i]}, a_2^{[i]}, \ldots, a_\ell^{[i]})^\top$. Therefore, the optimization problem (1) can be rewritten as:

$$\begin{aligned} \min_{\boldsymbol{w},\boldsymbol{\xi}} \quad & \textstyle\sum_{i=1}^{|\mathcal{I}|} \xi_i + C\mathbf{1}^\top \boldsymbol{w} \\ \text{s.t.} \quad & \boldsymbol{w} \succcurlyeq \mathbf{0};\ \mathbf{a}_i^\top \boldsymbol{w} \geq 1 - \xi_i,\ \xi_i \geq 0,\ \forall i, \end{aligned} \tag{2}$$

where $\mathbf{1}$ is the all-one column vector. To obtain the corresponding dual problem of (2), we write the Lagrangian as follows:

$$\begin{aligned} L &= C\mathbf{1}^\top \boldsymbol{w} - \mathbf{p}^\top \boldsymbol{w} + \textstyle\sum_{i=1}^{|\mathcal{I}|} u_i(1 - \xi_i - \mathbf{a}_i^\top \boldsymbol{w}) + \sum_{i=1}^{|\mathcal{I}|} \xi_i - \sum_{i=1}^{|\mathcal{I}|} \beta_i \xi_i \\ &= (C\mathbf{1} - \mathbf{p} - A\mathbf{u})^\top \boldsymbol{w} + \textstyle\sum_{i=1}^{|\mathcal{I}|} (1 - \beta_i - u_i)\xi_i + \mathbf{1}^\top \mathbf{u}, \end{aligned} \tag{3}$$

where $\mathbf{p} \succcurlyeq 0$, $\boldsymbol{\beta} \succcurlyeq 0$, and $\mathbf{u} \succcurlyeq 0$ are the Lagrange dual variables. We have introduced a matrix $A = (\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_{|\mathcal{I}|}) \in \mathcal{R}^{\ell \times |\mathcal{I}|}$. In order for L to have a finite infimum, $1 - \beta_i - u_i = 0, \forall i$, and $C\mathbf{1} - \mathbf{p} - A\mathbf{u} = 0$ must hold. This leads to $0 \leq u_i \leq 1, \forall i$, and $A\mathbf{u} \preccurlyeq C\mathbf{1}$. Putting these back into the Lagrangian, we obtain the dual problem as:

$$\begin{aligned} \max_{\mathbf{u}} \quad & \mathbf{1}^\top \mathbf{u} \\ \text{s.t.} \quad & A\mathbf{u} \preccurlyeq C\mathbf{1},\ \mathbf{0} \preccurlyeq \mathbf{u} \preccurlyeq \mathbf{1}. \end{aligned} \tag{4}$$

*Appearing in Proc. International Conference on Machine Learning, 2013.

2.2. Hashing with the l_∞ norm regularization

Using the l_∞ norm regularization The primal problem similar to the case of l_1 regularization (1) is formulated as:

$$\begin{aligned} \min_{\boldsymbol{w},\boldsymbol{\xi}} \quad & \textstyle\sum_{i=1}^{|\mathcal{I}|} \xi_i + C\|\boldsymbol{w}\|_\infty \\ \text{s.t.} \quad & \boldsymbol{w} \succcurlyeq \mathbf{0};\ d_{\mathcal{H}}(\mathbf{x}_i,\mathbf{x}_i^-) - d_{\mathcal{H}}(\mathbf{x}_i,\mathbf{x}_i^+) \geq 1-\xi_i,\ \xi_i \geq 0,\ \forall i. \end{aligned} \tag{5}$$

Clearly, the above optimization problem is equivalent to the following problem:

$$\begin{aligned} \min_{\boldsymbol{w},\boldsymbol{\xi}} \quad & \textstyle\sum_{i=1}^{|\mathcal{I}|} \xi_i \\ \text{s.t.} \quad & \boldsymbol{w} \succcurlyeq \mathbf{0},\ \|\boldsymbol{w}\|_\infty \leq C'; \\ & d_{\mathcal{H}}(\mathbf{x}_i,\mathbf{x}_i^-) - d_{\mathcal{H}}(\mathbf{x}_i,\mathbf{x}_i^+) \geq 1-\xi_i,\ \xi_i \geq 0,\ \forall i. \end{aligned} \tag{6}$$

Here C' is a positive control factor. Due to the fact that $\boldsymbol{w} \succcurlyeq \mathbf{0}$ and $\|\boldsymbol{w}\|_\infty \leq C'$, we have the relation: $\mathbf{0} \preccurlyeq \boldsymbol{w} \preccurlyeq C'$. Therefore, the optimization problem (6) becomes:

$$\begin{aligned} \min_{\boldsymbol{w},\boldsymbol{\xi}} \quad & \textstyle\sum_{i=1}^{|\mathcal{I}|} \xi_i \\ \text{s.t.} \quad & \mathbf{0} \preccurlyeq \boldsymbol{w} \preccurlyeq C'\mathbf{1};\ d_{\mathcal{H}}(\mathbf{x}_i,\mathbf{x}_i^-) - d_{\mathcal{H}}(\mathbf{x}_i,\mathbf{x}_i^+) \geq 1-\xi_i,\ \xi_i \geq 0,\ \forall i. \end{aligned} \tag{7}$$

The corresponding Lagrangian of (7) is formulated as:

$$\begin{aligned} L \ & = \mathbf{q}^\top(\boldsymbol{w} - C'\mathbf{1}) - \mathbf{p}^\top\boldsymbol{w} + \textstyle\sum_{i=1}^{|\mathcal{I}|} u_i(1-\xi_i-\mathbf{a}_i^\top\boldsymbol{w}) + \sum_{i=1}^{|\mathcal{I}|}\xi_i - \sum_{i=1}^{|\mathcal{I}|}\beta_i\xi_i \\ & = (\mathbf{q}-\mathbf{p}-A\mathbf{u})^\top\boldsymbol{w} - C'\mathbf{q}^\top\mathbf{1} + \textstyle\sum_{i=1}^{|\mathcal{I}|}(1-\beta_i-u_i)\xi_i + \mathbf{1}^\top\mathbf{u}, \end{aligned} \tag{8}$$

where $\mathbf{p},\mathbf{q},\mathbf{u}$ are Lagrangian multipliers. In order for L to have a finite infimum, $1-\beta_i-u_i = 0, \forall i$, and $\mathbf{q}-\mathbf{p}-A\mathbf{u} = 0$ must hold, leading to $0 \leq u_i \leq 1, \forall i$ and $A\mathbf{u} \preccurlyeq \mathbf{q}$. We then have the dual problem:

$$\begin{aligned} \max_{\mathbf{u},\mathbf{q}} \quad & \mathbf{1}^\top\mathbf{u} - C'\mathbf{1}^\top\mathbf{q} \\ \text{s.t.} \quad & A\mathbf{u} \preccurlyeq \mathbf{q},\ \mathbf{0} \preccurlyeq \mathbf{u} \preccurlyeq \mathbf{1}. \end{aligned} \tag{9}$$

The above optimization problem is equivalent to the following one:

$$\begin{aligned} \min_{\mathbf{u},\mathbf{q}} \quad & -\mathbf{1}^\top\mathbf{u} + C'\mathbf{1}^\top\mathbf{q} \\ \text{s.t.} \quad & A\mathbf{u} \preccurlyeq \mathbf{q},\ \mathbf{0} \preccurlyeq \mathbf{u} \preccurlyeq \mathbf{1}. \end{aligned} \tag{10}$$

3. Learning hash functions with general convex loss

3.1. Hashing with the l_1 norm regularization

Using the l_1 norm regularization to control the capacity, we may define the primal optimization problem as:

$$\min_{\boldsymbol{w},\boldsymbol{\rho}} \textstyle\sum_{i=1}^{|\mathcal{I}|} f(\rho_i) + C\|\boldsymbol{w}\|_1, \text{s.t. } \boldsymbol{w} \succcurlyeq \mathbf{0}; \rho_i = \mathbf{a}_i^\top\boldsymbol{w},\ \forall i. \tag{11}$$

We need to derive a meaningful Lagrange dual in order to use column generation. The Lagrange can be derived as:

$$\begin{aligned} L \ & = \textstyle\sum_{i=1}^{|\mathcal{I}|} f(\rho_i) + C\mathbf{1}^\top\boldsymbol{w} - \mathbf{p}^\top\boldsymbol{w} + \sum_{i=1}^{|\mathcal{I}|} u_i(\mathbf{a}_i^\top\boldsymbol{w} - \rho_i) \\ & = C\mathbf{1}^\top\boldsymbol{w} - \mathbf{p}^\top\boldsymbol{w} + \textstyle\sum_{i=1}^{|\mathcal{I}|} u_i(\mathbf{a}_i^\top\boldsymbol{w}) - (\mathbf{u}^\top\boldsymbol{\rho} - \sum_{i=1}^{|\mathcal{I}|} f(\rho_i)), \end{aligned} \tag{12}$$

where $\mathbf{p} \succcurlyeq \mathbf{0}$ and $\mathbf{u}$ are Lagrange multipliers. With the definition of Fenchel conjugate, we have the following relation: $\inf_{\boldsymbol{\rho}} L = -\sup_{\boldsymbol{\rho}} (\mathbf{u}^\top\boldsymbol{\rho} - \sum_{i=1}^{|\mathcal{I}|} f(\rho_i)) = -\sum_{i=1}^{|\mathcal{I}|} f^*(u_i)$ and in order to have a finite infimum, $C\mathbf{1} - \mathbf{p} + A\mathbf{u} = \mathbf{0}$. Since $\mathbf{p} \succcurlyeq 0$, $A\mathbf{u} \succcurlyeq -C\mathbf{1}$ must hold. Consequently, the dual problem can be expressed as:

$$\min_{\mathbf{u}} \quad \textstyle\sum_{i=1}^{|\mathcal{I}|} f^*(u_i), \text{s.t. } A\mathbf{u} \succcurlyeq -C\mathbf{1}. \tag{13}$$

Here $f^*(\cdot)$ is the Fenchel conjugate of $f(\cdot)$. By reversing the sign of $\mathbf{u}$, we can reformulate (13) as its equivalent form:

$$\begin{aligned} \min_{\mathbf{u}} \quad & \textstyle\sum_{i=1}^{|\mathcal{I}|} f^*(-u_i) \\ \text{s.t.} \quad & A\mathbf{u} \preccurlyeq C\mathbf{1}. \end{aligned} \tag{14}$$

From the Karush-Kuhn-Tucker (KKT) conditions between the primal (11) and the dual (14), the relation $u_i^\star = -f'(\rho_i^\star)$ holds at optimality, for a given smooth function $f(\cdot)$.

In the case of the squared hinge loss, $f(z) = [\max(0, 1-z)]^2$, the Fenchel duality $f^*(u_i)$ is computed as: $\sup_{\rho_i} [u_i\rho_i - f(\rho_i)] = \frac{u_i^2}{4} + u_i$. According to the KKT conditions, we have $u_i^\star = 2(1-\rho_i^\star)$ at optimality. As a result, we derive the following dual problem for the squared hinge loss case:

$$\begin{aligned} \min_{\mathbf{u}} \quad & \textstyle\sum_{i=1}^{|\mathcal{I}|} (\frac{u_i^2}{4} - u_i) \\ \text{s.t.} \quad & A\mathbf{u} \preccurlyeq C\mathbf{1}. \end{aligned} \tag{15}$$

As for the Logistic loss, $f(z) = \log(1+\exp(-z))$, the Fenchel duality $f^*(u_i)$ is calculated as: $\sup_{\rho_i} [u_i\rho_i - f(\rho_i)] = [(1+u_i)\log(1+u_i) - u_i\log(-u_i)]$. According to the KKT conditions, we have $u_i^\star = \frac{\exp(-\rho_i^\star)}{1+\exp(-\rho_i^\star)}$ at optimality. Therefore, the dual problem using the Logistic loss can be expressed as:

$$\begin{aligned} \min_{\mathbf{u}} \quad & \textstyle\sum_{i=1}^{|\mathcal{I}|} [(1-u_i)\log(1-u_i) + u_i\log(u_i)] \\ \text{s.t.} \quad & A\mathbf{u} \preccurlyeq C\mathbf{1}. \end{aligned} \tag{16}$$

3.2. Hashing with the l_∞ norm regularization

Using the l_∞ norm regularization, the primal problem can be written as:

$$\begin{aligned} \min_{\boldsymbol{w},\boldsymbol{\rho}} \quad & \textstyle\sum_{i=1}^{|\mathcal{I}|} f(\rho_i) + C\|\boldsymbol{w}\|_\infty \\ \text{s.t.} \quad & \boldsymbol{w} \succcurlyeq \mathbf{0}; \rho_i = \mathbf{a}_i^\top \boldsymbol{w},\ \forall i, \end{aligned} \tag{17}$$

where $f(\cdot)$ is a convex loss function. The above optimization problem is equivalent to:

$$\begin{aligned} \min_{\boldsymbol{w},\boldsymbol{\rho}} \quad & \textstyle\sum_{i=1}^{|\mathcal{I}|} f(\rho_i) \\ \text{s.t.} \quad & \|\boldsymbol{w}\|_\infty \le C',\ \boldsymbol{w} \succcurlyeq \mathbf{0};\ \rho_i = \mathbf{a}_i^\top \boldsymbol{w},\ \forall i, \end{aligned} \tag{18}$$

where C' is a positive control factor. Due to $\boldsymbol{w} \succcurlyeq \mathbf{0}$ and $\|\boldsymbol{w}\|_\infty \le C'$, we obtain $\mathbf{0} \preccurlyeq \boldsymbol{w} \preccurlyeq C'\mathbf{1}$. Therefore, we rewrite the optimization problem (18) as:

$$\begin{aligned} \min_{\boldsymbol{w},\boldsymbol{\rho}} \quad & \textstyle\sum_{i=1}^{|\mathcal{I}|} f(\rho_i) \\ \text{s.t.} \quad & \mathbf{0} \preccurlyeq \boldsymbol{w} \preccurlyeq C'\mathbf{1};\ \rho_i = \mathbf{a}_i^\top \boldsymbol{w},\ \forall i, \end{aligned} \tag{19}$$

The Lagrange can be derived as:

$$L = \sum_{i=1}^{|\mathcal{I}|} f(\rho_i) + \mathbf{q}^\top \boldsymbol{w} - C'\mathbf{q}^\top \mathbf{1} - \mathbf{p}^\top \boldsymbol{w} + \sum_{i=1}^{|\mathcal{I}|} u_i(\mathbf{a}_i^\top \boldsymbol{w} - \rho_i), \tag{20}$$

where $\mathbf{p}$, $\mathbf{q}$, $\mathbf{u}$ are Lagrange multipliers.

We take the first derivative of L w.r.t. $\boldsymbol{w}$ and set it to zero:

$$\frac{\partial L}{\partial \boldsymbol{w}} = \mathbf{q} - \mathbf{p} + \sum_{i=1}^{|\mathcal{I}|} u_i\mathbf{a}_i = \mathbf{q} - \mathbf{p} + A\mathbf{u} = 0. \tag{21}$$

Since $\mathbf{p} \succcurlyeq 0$, we have the relation $A\mathbf{u} \succcurlyeq -\mathbf{q}$. Moreover, we can derive the following relation:

$$\inf_{\mathbf{u},\boldsymbol{\rho}} L = -\sup_{\boldsymbol{\rho}} (\mathbf{u}^\top \boldsymbol{\rho} - \sum_{i=1}^{|\mathcal{I}|} f(\rho_i)) = -\sum_{i=1}^{|\mathcal{I}|} f^*(u_i) \tag{22}$$

Variable	$h(\mathbf{x}_i)$	$h(\mathbf{x}_i^-)$	$h(\mathbf{x}_i^+)$	$\lvert\tau_{i-}\rvert - \lvert\tau_{i+}\rvert$	$\tau_{i-}^2 - \tau_{i+}^2$
Case 1	1	1	1	0	0
Case 2	1	-1	1	2	4
Case 3	1	1	-1	-2	-4
Case 4	1	-1	-1	0	0
Case 5	-1	1	1	0	0
Case 6	-1	-1	1	-2	-4
Case 7	-1	1	-1	2	4
Case 8	-1	-1	-1	0	0

Table 1: Illustration of all possible cases for $|\tau_{i-}| - |\tau_{i+}|$ and $\tau_{i-}^2 - \tau_{i+}^2$.

where $f^*(\cdot)$ is the Fenchel conjugate of $f(\cdot)$. Therefore, the dual problem is formulated as:

$$\begin{array}{ll} \min\limits_{\mathbf{u},\mathbf{q}} & \sum_{i=1}^{|\mathcal{I}|} f^*(u_i) + C'\mathbf{q}^\top \mathbf{1} \\ \text{s.t.} & A\mathbf{u} \succcurlyeq -\mathbf{q}. \end{array} \tag{23}$$

By reversing the sign of $\mathbf{u}$, we can reformulate (23) as its equivalent form:

$$\begin{array}{ll} \min\limits_{\mathbf{u},\mathbf{q}} & \sum_{i=1}^{|\mathcal{I}|} f^*(-u_i) + C'\mathbf{q}^\top \mathbf{1} \\ \text{s.t.} & A\mathbf{u} \preccurlyeq \mathbf{q}. \end{array} \tag{24}$$

Similar to (15), the dual problem using the squared hinge loss is formulated as:

$$\begin{array}{ll} \min\limits_{\mathbf{u},\mathbf{q}} & \sum_{i=1}^{|\mathcal{I}|} (\frac{u_i^2}{4} - u_i) + C'\mathbf{q}^\top \mathbf{1} \\ \text{s.t.} & A\mathbf{u} \preccurlyeq \mathbf{q}. \end{array} \tag{25}$$

Similar to (16), the dual problem using the logistic loss is written as:

$$\begin{array}{ll} \min\limits_{\mathbf{u},\mathbf{q}} & \sum_{i=1}^{|\mathcal{I}|} [(1-u_i)\log(1-u_i) + u_i \log(u_i)] + C'\mathbf{q}^\top \mathbf{1} \\ \text{s.t.} & A\mathbf{u} \preccurlyeq \mathbf{q}. \end{array} \tag{26}$$

4. Finding the most violated dual constraint

The core idea of column generation is to generate a small subset of variables, each of which is sequentially found by selecting the most violated dual constraints in the dual optimization problem. This process is equivalent to inserting several primal variables into the primal optimization problem. Here, the subproblem for generating the most violated dual constraint (i.e., to find the best hash function) can be defined as:

$$h^\star(\cdot) = \mathop{\arg\max}_{h(\cdot)\in\mathcal{H}} \sum_{i=1}^{|\mathcal{I}|} u_i a^{[i]} = \mathop{\arg\max}_{h(\cdot)\in\mathcal{H}} \sum_{i=1}^{|\mathcal{I}|} u_i (|h(\mathbf{x}_i) - h(\mathbf{x}_i^-)| - |h(\mathbf{x}_i) - h(\mathbf{x}_i^+)|). \tag{27}$$

In order to obtain a smoothly differentiable objective function, we transform (27) to the following equivalent form:

$$h^\star(\cdot) = \mathop{\arg\max}_{h(\cdot)\in\mathcal{H}} \sum_{i=1}^{|\mathcal{I}|} u_i [(h(\mathbf{x}_i) - h(\mathbf{x}_i^-))^2 - (h(\mathbf{x}_i) - h(\mathbf{x}_i^+))^2] \tag{28}$$

The proof of showing the equivalence between (27) and (28) is given as follows. For notational simplicity, let τ_{i+} and τ_{i-} denote $h(\mathbf{x}_i) - h(\mathbf{x}_i^+)$ and $h(\mathbf{x}_i) - h(\mathbf{x}_i^-)$, respectively. Therefore, the optimization problem (27) can be simplified as:

$$h^\star(\cdot) = \mathop{\arg\max}_{h(\cdot)\in\mathcal{H}} \sum_{i=1}^{|\mathcal{I}|} u_i (|\tau_{i-}| - |\tau_{i+}|) \tag{29}$$

Dataset	MNIST	USPS	LABELME [4, 5]	SCENE-15 [2]	ISOLET	PASCAL07 [1]
Size	70,000	9,298	50,000	4,485	7,797	9,963
Dimension	784	256	512	6,200	617	2712
Classes	10	10	12	15	26	20

Table 2: Summary of the 6 datasets used in the experiments.

Correspondingly, the optimization problem (28) can be rewritten as:

$$h^{\star}(\cdot) = \underset{h(\cdot)\in\mathcal{H}}{\arg\max} \sum_{i=1}^{|\mathcal{I}|} u_i [\tau_{i-}^2 - \tau_{i+}^2]. \tag{30}$$

Since $h(\mathbf{x}_i), h(\mathbf{x}_i^+), h(\mathbf{x}_i^-) \in \{-1, 1\}$, we obtain the following relation such that $|\tau_{i+}|, |\tau_{i-}| \in \{0, 2\}$ and $\tau_{i+}^2, \tau_{i-}^2 \in \{0, 4\}$. As a result, we investigate all the possible cases for $|\tau_{i-}| - |\tau_{i+}|$ and $\tau_{i-}^2 - \tau_{i+}^2$ in Table 1. From Table 1, it is seen that $\tau_{i-}^2 - \tau_{i+}^2 = 2(|\tau_{i-}| - |\tau_{i+}|)$ in all possible cases, leading to the equivalence between (27) and (28).

5. Dataset description

Table 2 gives a summary of the 6 six datasets used in the experiments. More specifically, the USPS dataset consists of 9,298 handwritten images, each of which is resized to 16×16 pixels. This dataset is split into two subsets at random (70% for training and 30% for testing).

The MNIST dataset is composed of 70,000 images of handwritten digits and is thus divided into 10 classes of 28×28 pixel images. This dataset is randomly partitioned into a training subset (66,000 images) and a testing subset (4,000 images). We select 2000 images from the training subset to generate a set of triplets used for learning hash functions. In the above two handwritten image datasets, the original gray-scale intensity values of each image are used as features for image representation.

The ISOLET dataset contains 7,797 recordings of 150 subjects speaking the 26 letters of the alphabet. Each subject spoke each letter twice. This dataset is randomly divided into a training subset (5,459 spoken letters) and a testing subset (2,338 spoken letters). Each letter is represented as a 617-dimensional feature vector.

The SCENE-15 dataset totally constitutes 4,485 images of 9 outdoor scenes and 6 indoor scenes. Each image is divided into 31 sub-windows, each of which is represented as a histogram of 200 visual code words. A concatenation of the histograms associated with 31 sub-windows is used to represent an image, resulting in a 6,200-dimensional feature vector. This dataset is randomly divided into a training subset (1500 samples) and a testing subset (2985 samples).

The LABELME dataset is a subset of the original LabelMe dataset, and consists of 50,000 images that are categorized into 12 classes. Each image is 256×256 pixels, and then represented as a 512-dimensional gist feature.

The PASCAL07 dataset is a subset of the PASCAL VOC 2007 dataset, and contains 9,963 images of 20 object classes. Each image is represented by a 2712-dimensional feature vector, as was the case in `http://lear.inrialpes.fr/people/guillaumin/data.php`. This dataset is randomly separated into a training subset (70%) and a testing subset (30%). The training/testing split is repeated 5 times, and the average performance over these 5 trials is reported in the experiments.

References

[1] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The PASCAL visual object classes challenge 2007 results. `http://www.pascal-network.org/challenges/VOC/voc2007/workshop/index.html`.

[2] S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *Proc. IEEE Conf. Computer Vision & Pattern Recognition*, volume 2, pages 2169–2178, 2006.

[3] X. Li, G. Li, C. Shen, A. van den Hengel, and A. Dick. Learning hash functions using column generation. In *Proc. Int. Conf. Machine Learning*, 2013.

[4] B. Russell, A. Torralba, K. Murphy, and W. Freeman. LabelMe: a database and web-based tool for image annotation. *Int. J. Computer Vision*, 77(1):157–173, 2008.

[5] R. Uetz and S. Behnke. Large-scale object recognition with CUDA-accelerated hierarchical neural networks. In *IEEE Int. Conf. Intelligent Computing & Intelligent Systems*, volume 1, pages 536–541, 2009.