Department IV Computer Science and Electrical Engineering
Technische Universität Berlin

Machine Learning II

Exercise Sheet 1

Due Date:

Tuesday, Mai 03, 2016

Group: BTMNLERR
Mario Johannes Neumann 327468
Leon Eifler 331653
Borbala Tasnadi
Ronnie Rampersad

May 3, 2016

Exercise 1

We want to show scaling, translation and rotation invariancies for the error term, which is given as

$$E(w) = \sum_i |X_i - \sum_j w_{ij} X_j|. \tag{1}$$

Ex 1.a - sclaing invariance

We scale each of the X_i by a fixed scalar $alpha \in \mathbb{R}^+ \setminus \{0\}$. Then we get:

$$\sum_i |\alpha X_i - \sum_j w_{ij} \alpha X_j| = \sum_i |\alpha(X_i - \sum_j w_{ij} X_j)| = |\alpha| \sum_i |X_i - \sum_j w_{ij} X_j|$$

So the scaled error term is the same as the original error multiplied by a constant, positive number α. Thus it has the same minimum as the original term.

Ex 1.b - translation invariance

Now we translate each of the X_i by the same vector v. Here we will use that $\sum_j w_{ij} = 1$. Then we can show:

$$\sum_i |X_i + v - \sum_j w_{ij}(X_j + v)|. = \sum_i |X_i + v - \sum_j w_{ij} X_j - \sum_j w_{ij} v)|. = \sum_i |X_i + v - \sum_j w_{ij} X_j - v \sum_j w_{ij})|$$

So the translated terms cancel each other out and the error is exactly the same as the old error and thus has the same minimum.

Ex 1.c - rotation invariance

Let $U \in \mathbb{R}^{n \times n}$ be an orthogonal matrix. Keep in mind that orthogonal matrizes preserve distances, i.e. $|Ux| = |x|$ for any $x \in \mathbb{R}^n$.

We will also use that matrix-multiplikation is a linear operation. Then we get for our error-term the following.

$$\sum_i |UX_i - \sum_j w_{ij} UX_j|. = \sum_i |U(X_i - \sum_j w_{ij} X_j)|. = \sum_i |X_i - \sum_j w_{ij} X_j|.$$

This is again the original error and thus the minimum is preserved.

Havin proven all of the above, we have also proven invariance w.r.t. arbitraty rotations. If we wan to rotate around an arbitraty fixed point that is not the origin, simply translate the whole thing so that the rotation centre is the origin, do the rotation and then translate back. As all of these operations are invariant, their sequential use is also invariant.

Exercise 2

(i) The matrix C is defined as follows.

$$
\begin{aligned}
C &= (\mathbb{1} x^T - \eta) \cdot (\mathbb{1} x^T - \eta)^T & (2)\\
&= \left[\begin{pmatrix} x^T \\ \vdots \\ x^T \end{pmatrix} - \begin{pmatrix} \eta_1^T \\ \vdots \\ \eta_k^T \end{pmatrix} \right] \cdot \left[\begin{pmatrix} x^T \\ \vdots \\ x^T \end{pmatrix} - \begin{pmatrix} \eta_1^T \\ \vdots \\ \eta_k^T \end{pmatrix} \right]^T & (3)\\
&= \begin{bmatrix} x^T - \eta_1^T \\ \vdots \\ x^T - \eta_k^T \end{bmatrix} \cdot \begin{bmatrix} x^T - \eta_1^T \\ \vdots \\ x^T - \eta_k^T \end{bmatrix}^T & (4)\\
&= \begin{bmatrix} (x - \eta_1)^T \\ \vdots \\ (x - \eta_k)^T \end{bmatrix} \cdot \begin{bmatrix} x - \eta_1 \dots x - \eta_k \end{bmatrix} & (5)
\end{aligned}
$$

So the i-th entry in the j-th row of C is
$C_{ij} = (x^T - \eta_i^T)(x - \eta_j) = (x - \eta_i)^T (x - \eta_j) = < x - \eta_i, x - \eta_j >$.
Assume that the entries of w add up to 1. Then, for the error ε, we compute

$$
\begin{aligned}
\varepsilon = \left| x - \sum_{j=1}^k w_j \eta_j \right|^2 = \left| \big(\underbrace{\sum_{j=1}^k w_j}_{=1} \big) x - \sum_{j=1}^k w_j \eta_j \right|^2 &= \left| \sum_{j=1}^k w_j (x - \eta_j) \right|^2 & (6)\\
&= < \sum_{i=1}^k w_i (x - \eta_i), \sum_{j=1}^k w_j (x - \eta_j) > & (7)\\
&= \sum_{i,j=1}^k w_i w_j < x - \eta_i, x - \eta_j > \quad \text{by bilinearity} & (8)\\
&= \sum_{i,j=1}^k w_i w_j C_{ij} & (9)\\
&= w^T C w \ . & (10)
\end{aligned}
$$

Hence, minimizing ε is equivalent to minimizing $w^T C w$ under the condition that the entries of w add up to 1.

(iii) Consider the Lagrange function
$\mathcal{L}(w, \lambda) := w^T C w - \lambda \left(\sum_{j=1}^k w_j - 1 \right)$

where λ is the Langrange multiplier. If $w^T C w$ has a minimum over the set of weight vectors w satisfying

$$\sum_{j=1}^{k} w_j = 1$$

then there exists a λ such that the partial derivatives of $\mathcal{L}$:

$$(1) \quad \frac{\partial}{\partial w}\mathcal{L}(w,\lambda) = 2Cw - \nVdash\lambda \tag{11}$$

$$(2) \quad \frac{\partial}{\partial \lambda}\mathcal{L}(w,\lambda) = -\sum_{j=1}^{k} w_j + 1 \tag{12}$$

are zero. Now assume that $\overline{w}$ solves the equation $C\overline{w} = \nVdash$. and that the sum of the entries of $\overline{w}$ is nonzero. Then the weight vector $w := \frac{\overline{w}}{\sum_{j=1}^{k} w_j}$ solves (1), for $\lambda = \frac{2}{\sum_{j=1}^{k} w_j}$.

(ii) Let $\overline{w}$ be as written above. In case C is regular, $\overline{w}$ is just $C^{-1}\nVdash$ and it satisfies

$\sum_{j=1}^{k} \overline{w}_j = \nVdash^T \overline{w} = \nVdash^T C^{-1} \nVdash$.

By part (iii), the optimal weight vector is then given by

$w = \frac{\overline{w}}{\sum_{j=1}^{k} w_j} = \frac{C^{-1}\nVdash}{\nVdash^T C^{-1}\nVdash}$.