Supplement for Paper 1738

Bo Liu Nuno Vasconcelos
University of California, San Diego
La Jolla, CA 92093
boliu@ucsd.edu, nvasconcelos@ucsd.edu

A. Derivation of updated parameter posterior

Since $p(\mathbf{w}|X, \mathbf{y})$ is not a zero mean distribution, we consider the transformation

$$\mathbf{w}' = \mathbf{w} - \mu_{\mathbf{w}} \tag{1.1}$$

such that

$$p(\mathbf{w}'|X, \mathbf{y}) \quad = \quad G(\mathbf{w}', \mathbf{0}, \Sigma_{\mathbf{w}}), \tag{1.2}$$

with $\Sigma_{\mathbf{w}}$ given by (5). The regression model of (1) can then be written as

$$y = g(\mathbf{x}) + \phi(\mathbf{x})^T \mu_{\mathbf{w}} + \epsilon \qquad g(\mathbf{x}) = \phi(\mathbf{x})^T \mathbf{w}'. \tag{1.3}$$

Introducing the change of variables

$$z = y - \phi(\mathbf{x})^T \mu_{\mathbf{w}} \tag{1.4}$$

leads to a process identical to (1), i.e.

$$z = g(\mathbf{x}) + \epsilon \qquad g(\mathbf{x}) = \phi(\mathbf{x})^T \mathbf{w}' \tag{1.5}$$

with zero-mean prior $p(\mathbf{w}'|X, z) = G(\mathbf{w}', \mathbf{0}, \Sigma_{\mathbf{w}})$ and iid Gaussian noise $\epsilon \sim \mathcal{N}(0, \sigma_n^2)$. Using (20) and (21), it follows from (15), (1.1), and (1.4) that

$$p(\mathbf{w}'|X^+, \mathbf{z}^+, X, \mathbf{y}) \propto p(\mathbf{z}^+|X^+, \mathbf{w}')p(\mathbf{w}'|X, \mathbf{y}). \tag{1.6}$$

Similarly to (2)-(5), this leads to a parameter posterior

$$p(\mathbf{w}'|X^+, \mathbf{z}^+, X, \mathbf{y}) = G(\mathbf{w}', \mu_{\mathbf{w}'}, \Sigma_{\mathbf{w}'}) \tag{1.7}$$

of hyperparameters

$$\mu_{\mathbf{w}'} = \frac{1}{\sigma_n^2} \Gamma_{\mathbf{w}}^{-1} \Phi^+ \mathbf{z}^+ \qquad \Sigma_{\mathbf{w}'} = \Gamma_{\mathbf{w}}^{-1}, \tag{1.8}$$

with $\Gamma_{\mathbf{w}}$ as in (19). Using (1.1) finally leads to (17)-(18).

B. Derivation of the predictive distribution

By combining (1.7) with the procedure used to derive (6) from (2), it follows that the updated predictive distribution is

$$p(z_*^+|\mathbf{x}_*^+, X^+, \mathbf{z}^+, X, \mathbf{y}) = G(z_*^+, \mu'(\mathbf{x}_*^+), \sigma'(\mathbf{x}_*^+)) \tag{2.9}$$

with parameters

$$\begin{aligned} \mu'(\mathbf{x}_*^+) \quad &= \quad {\phi_*^+}^T \mu_{\mathbf{w}'} & (2.10) \\ \sigma'(\mathbf{x}_*^+) \quad &= \quad \sigma_n^2 + {\phi_*^+}^T \Gamma_{\mathbf{w}}^{-1} \phi_*^+, & (2.11) \end{aligned}$$

where $\phi_*^+ = \phi(\mathbf{x}_*^+)$. The predictive distribution of (22)-(23) then follows from (1.4), with

$$\sigma^+(\mathbf{x}_*^+) = \sigma'(\mathbf{x}_*^+) \tag{2.12}$$

and

$$\begin{aligned} \mu^+(\mathbf{x}_*^+) &= {\phi_*^+}^T \mu_{\mathbf{w}} + \mu'(\mathbf{x}_*^+) & (2.13) \\ &= \quad {\phi_*^+}^T \mu_{\mathbf{w}} + \frac{1}{\sigma_n^2} {\phi_*^+}^T \Gamma_{\mathbf{w}}^{-1} \Phi^+ (\mathbf{y}^+ - {\Phi^+}^T \mu_{\mathbf{w}}), \end{aligned}$$

where we have also used (21) and (18). Since, from the matrix inversion lemma [1]

$$\Gamma_{\mathbf{w}}^{-1} = \Sigma_{\mathbf{w}} - \Sigma_{\mathbf{w}} \Phi^+ \Omega_{\mathbf{w}}^{-1} {\Phi^+}^T \Sigma_{\mathbf{w}} \tag{2.14}$$

with $\Omega_{\mathbf{w}} = {\Phi^+}^T \Sigma_{\mathbf{w}} \Phi^+ + \sigma_n^2 I$, it follows that

$$\begin{aligned} \Gamma_{\mathbf{w}}^{-1} \Phi^+ \quad &= \quad \Sigma_{\mathbf{w}} \Phi^+ [I - \Omega_{\mathbf{w}}^{-1} {\Phi^+}^T \Sigma_{\mathbf{w}} \Phi^+] \\ &= \quad \Sigma_{\mathbf{w}} \Phi^+ \Omega_{\mathbf{w}}^{-1} \sigma_n^2 \end{aligned}$$

and the predictive distribution of (22)-(23) has parameters

$$\begin{aligned} \mu^+(\mathbf{x}_*^+) \quad &= \quad {\phi_*^+}^T \mu_{\mathbf{w}} + {\phi_*^+}^T \Sigma_{\mathbf{w}} \Phi^+ \Omega_{\mathbf{w}}^{-1} (\mathbf{y}^+ - {\Phi^+}^T \mu_{\mathbf{w}}) \\ &= \quad \mu(\mathbf{x}_*^+) + {\phi_*^+}^T \Sigma_{\mathbf{w}} \Phi^+ \Omega_{\mathbf{w}}^{-1} (\mathbf{y}^+ - \mu(\mathbf{X}^+)) \\ &= \quad \mu(\mathbf{x}_*^+) + {\mathbf{k}_*^+}^T [K^+ + \sigma_n^2 I]^{-1} \mathbf{e} & (2.15) \\ \sigma^+(\mathbf{x}_*^+) \quad &= \quad \sigma_n^2 + {\phi_*^+}^T \Sigma_{\mathbf{w}} \phi_*^+ \\ &- \quad {\phi_*^+}^T \Sigma_{\mathbf{w}} \Phi^+ \Omega_{\mathbf{w}}^{-1} {\Phi^+}^T \Sigma_{\mathbf{w}} \phi_*^+ \\ &= \quad \sigma(\mathbf{x}_*^+) - {\phi_*^+}^T \Sigma_{\mathbf{w}} \Phi^+ \Omega_{\mathbf{w}}^{-1} {\Phi^+}^T \Sigma_{\mathbf{w}} \phi_*^+ \\ &= \quad \sigma(\mathbf{x}_*^+) - {\mathbf{k}_*^+}^T [K^+ + \sigma_n^2 I]^{-1} \mathbf{k}_*^+ & (2.16) \end{aligned}$$

Table 1. MAE of various methods on the proposed dataset. As NTA and NTF are two no transfer modes, they don't have confident limits for each view.

Heavy	V1-1	V1-2	V1-3	V2-1	V2-2	V2-3	V3-1
GPA	2.78 ± 0.80	3.09 ± 1.03	1.47 ± 0.75	1.73 ± 0.68	2.39 ± 0.78	1.39 ± 0.57	2.14 ± 0.54
GPTL	3.36 ± 3.82	3.31 ± 6.64	1.47 ± 1.22	1.59 ± 0.79	2.81 ± 0.80	1.36 ± 0.54	2.27 ± 0.77
HGP	5.00 ± 0.63	5.60 ± 1.12	2.33 ± 0.71	2.92 ± 0.83	3.26 ± 1.40	2.29 ± 1.03	3.00 ± 0.51
WKT	3.30 ± 0.62	6.58 ± 1.57	2.95 ± 2.10	2.44 ± 0.89	9.19 ± 4.93	2.68 ± 0.73	3.87 ± 1.05
FA	5.65 ± 3.62	13.75 ± 6.59	3.18 ± 1.56	2.99 ± 1.19	6.47 ± 1.99	2.32 ± 0.78	2.96 ± 0.98
NTA	4.23	6.51	**1.34**	2.32	21.55	3.97	3.79
NA	18.33 ± 5.83	28.08 ± 7.79	7.76 ± 4.67	8.46 ± 9.09	10.19 ± 6.32	7.14 ± 8.35	9.78 ± 4.88
NTF	1.89	1.64	0.89	1.07	1.67	0.85	1.07
V3-2	V3-3	V4-1	V4-2	V4-3	V5-1	V5-2	V5-3
2.30 ± 0.71	2.26 ± 0.60	1.86 ± 0.96	2.02 ± 0.66	2.01 ± 0.63	2.15 ± 1.05	2.18 ± 0.82	2.10 ± 0.71
3.51 ± 1.71	2.53 ± 0.84	1.93 ± 1.13	2.13 ± 2.57	2.38 ± 1.24	2.25 ± 3.76	2.53 ± 1.43	3.96 ± 0.80
3.51 ± 1.01	3.03 ± 0.59	2.41 ± 0.82	2.81 ± 1.23	3.20 ± 1.62	2.59 ± 1.35	2.93 ± 1.11	3.33 ± 1.11
6.15 ± 0.76	8.59 ± 3.51	2.78 ± 1.04	2.55 ± 0.53	3.06 ± 0.92	1.94 ± 1.09	2.67 ± 1.28	1.80 ± 1.28
5.66 ± 3.02	3.07 ± 1.06	2.14 ± 1.34	2.80 ± 1.56	2.70 ± 0.83	2.52 ± 2.24	3.03 ± 1.87	2.76 ± 1.06
6.47	11.05	2.29	3.20	3.68	**1.37**	**2.05**	3.17
14.51 ± 5.99	8.41 ± 4.40	8.08 ± 8.46	7.62 ± 7.85	7.40 ± 7.79	6.88 ± 5.56	7.47 ± 7.63	6.77 ± 6.77
1.23	1.20	1.27	2.02	2.16	1.38	2.42	2.64
V6-1	V6-2	V6-3	V7-1	V7-2	V7-3	V8-1	V8-2
1.99 ± 1.11	2.14 ± 0.91	2.30 ± 0.70	2.04 ± 0.78	2.16 ± 0.97	2.00 ± 0.58	2.24 ± 0.51	2.05 ± 0.80
1.61 ± 5.57	2.32 ± 1.24	2.84 ± 2.16	2.15 ± 1.60	2.12 ± 3.62	2.23 ± 1.04	2.73 ± 1.52	2.07 ± 5.49
2.60 ± 1.40	2.54 ± 0.88	3.31 ± 1.53	2.49 ± 0.90	3.69 ± 1.56	2.77 ± 0.86	3.70 ± 0.73	2.48 ± 0.56
1.68 ± 0.50	2.77 ± 0.46	3.06 ± 0.74	2.52 ± 1.10	3.20 ± 2.08	2.66 ± 0.65	3.42 ± 1.09	4.12 ± 0.98
1.99 ± 1.01	2.92 ± 1.52	3.18 ± 1.22	2.67 ± 0.67	2.81 ± 1.30	2.88 ± 0.69	2.63 ± 1.21	4.27 ± 1.53
1.41	3.20	3.10	**2.03**	2.14	2.57	3.86	3.78
7.34 ± 6.23	7.76 ± 7.66	7.96 ± 9.06	6.99 ± 4.86	8.44 ± 9.08	8.53 ± 10.33	11.29 ± 11.38	7.20 ± 5.42
1.34	2.37	3.39	1.64	1.76	1.63	2.86	3.43
V8-3	V9-1	V9-2	V9-3	**Avg.**	Time		
2.10 ± 0.43	1.78 ± 0.72	2.14 ± 0.71	1.99 ± 0.66	2.10 ± 0.82	0		
2.76 ± 1.42	1.77 ± 5.84	2.46 ± 1.21	2.14 ± 0.91	2.39 ± 2.95	0.72		
2.45 ± 0.58	2.48 ± 1.26	3.09 ± 0.92	3.15 ± 1.34	3.07 ± 1.29	0.01		
3.32 ± 0.69	1.27 ± 0.78	2.71 ± 1.38	2.41 ± 0.96	3.47 ± 2.43	0.86		
5.28 ± 3.56	2.63 ± 2.14	3.12 ± 1.95	2.68 ± 1.02	3.67 ± 3.10	3.32		
3.30	1.45	**1.94**	2.31	4.00 ± 4.05	0.11		
12.77 ± 7.93	7.54 ± 7.66	11.63 ± 10.46	10.62 ± 12.02	9.81 ± 8.77	0.05		
2.50	1.56	2.28	1.72	1.85 ± 0.70	1.00		

Weak	V1-1	V1-2	V1-3	V2-1	V2-2	V2-3	V3-1
GPA	3.62 ± 1.22	5.21 ± 1.70	2.20 ± 1.39	3.10 ± 1.46	5.30 ± 2.35	2.86 ± 1.79	2.62 ± 0.86
GPTL	3.91 ± 2.41	6.02 ± 4.43	3.27 ± 2.43	3.05 ± 2.09	7.43 ± 5.90	2.84 ± 2.90	4.62 ± 1.72
HGP	7.14 ± 2.51	19.33 ± 12.57	3.47 ± 1.53	5.17 ± 2.14	8.53 ± 6.67	3.64 ± 2.05	5.14 ± 2.05
WKT	4.69 ± 1.25	11.62 ± 3.69	3.76 ± 2.24	3.26 ± 1.14	7.37 ± 4.23	2.91 ± 2.47	3.81 ± 0.88
FA	3.23 ± 3.47	14.70 ± 6.41	4.27 ± 2.42	3.07 ± 1.32	6.62 ± 1.98	3.31 ± 1.23	3.30 ± 1.11
NTA	4.99	**3.83**	**2.12**	3.62	9.39	**2.70**	5.94
V3-2	V3-3	V4-1	V4-2	V4-3	V5-1	V5-2	V5-3
5.12 ± 1.62	2.86 ± 0.74	2.38 ± 1.52	2.94 ± 1.19	2.71 ± 1.16	3.21 ± 1.68	2.88 ± 1.54	2.65 ± 1.06
3.53 ± 3.03	4.59 ± 2.83	3.92 ± 1.95	2.64 ± 6.62	3.62 ± 2.34	4.27 ± 2.55	3.18 ± 6.38	3.83 ± 2.63
5.71 ± 1.59	4.33 ± 1.21	2.70 ± 1.49	4.11 ± 2.08	3.40 ± 1.91	3.61 ± 2.97	5.41 ± 4.01	3.15 ± 1.93
4.94 ± 1.74	7.18 ± 3.83	2.78 ± 1.73	7.43 ± 7.46	3.55 ± 1.83	2.88 ± 1.80	4.66 ± 4.70	3.90 ± 1.99
4.45 ± 2.99	3.71 ± 2.03	2.41 ± 1.64	3.26 ± 1.69	3.11 ± 0.96	3.03 ± 2.53	3.74 ± 2.38	3.19 ± 1.63
7.15	**2.64**	2.39	4.84	5.35	3.20	3.85	5.78
V6-1	V6-2	V6-3	V7-1	V7-2	V7-3	V8-1	V8-2
2.46 ± 1.55	3.15 ± 1.65	3.17 ± 1.46	2.89 ± 1.14	2.81 ± 1.39	3.10 ± 1.22	2.48 ± 0.69	3.68 ± 1.34
4.68 ± 2.61	3.35 ± 8.23	3.65 ± 2.40	3.71 ± 2.64	3.46 ± 3.50	3.50 ± 2.38	4.65 ± 3.32	3.96 ± 2.48
2.91 ± 2.44	5.27 ± 2.04	6.02 ± 4.38	3.13 ± 2.10	3.56 ± 1.71	4.67 ± 0.94	4.90 ± 1.39	3.88 ± 0.95
2.56 ± 2.44	10.20 ± 8.95	3.83 ± 1.73	3.79 ± 2.09	3.43 ± 2.09	3.40 ± 1.01	4.37 ± 1.46	4.06 ± 1.34
2.24 ± 1.34	3.58 ± 1.71	3.64 ± 1.60	2.78 ± 1.32	2.96 ± 1.20	3.20 ± 0.84	3.61 ± 1.86	4.09 ± 1.58
3.08	4.91	5.51	2.89	4.40	4.23	4.37	4.57
V8-3	V9-1	V9-2	V9-3	**Avg.**	Time		
2.95 ± 0.76	3.07 ± 1.65	3.13 ± 1.40	3.20 ± 1.64	3.18 ± 1.61	0		
4.13 ± 1.38	3.54 ± 2.92	4.62 ± 4.98	3.65 ± 5.33	3.99 ± 4.06	0.64		
7.59 ± 5.07	5.01 ± 6.22	4.51 ± 2.29	5.14 ± 1.56	5.24 ± 4.81	0.01		
2.78 ± 0.65	3.18 ± 2.19	6.01 ± 5.26	3.84 ± 2.71	4.67 ± 3.57	0.53		
5.01 ± 5.50	2.63 ± 2.35	3.64 ± 2.33	3.24 ± 1.70	3.93 ± 3.35	2.83		
5.27	2.85	**2.75**	4.38	4.33 ± 1.61	0.12		

where we have used (7)-(8) and $\mathbf{e}, \mu(\mathbf{X}^+), \mathbf{k}_*^+$, and K^+ are as given in (25)-(28).

References

[1] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, 2006. 1