Sparse Variational Inference for Generalized Gaussian Process Models

Table 1. Summary of data sets. Values in parentheses refer to number of categories

NAME	SAMPLES	NO. DIM.	MODEL TYPE
UCSDPEDS1L	4000	30	COUNT
ABALONE	4177	8	COUNT
BIKEHOUR	8734	12	COUNT
YEAST	1484	8	BINARY
USPS35	1540	256	BINARY
MUSK	6958	166	BINARY
STOCK (5)	950	9	ORDINAL
BANK (10)	8192	32	ORDINAL
MOVIELENS (5)	4662	6	ORDINAL

1. Supplementary Figures

The datasets used in the experiment are summarized in Table 1. The dataset *ucsdpeds1l* contains counts of pedestrians extracted from video data and was used in (Chan & Vasconcelos, 2012). The *bikehour* dataset contains casual and registered bicycle rental counts over 2011 in Washington D.C. (Fanaee-T & Gama, 2014). The datasets *stock* and *bank* were used in previous ordinal regression experiments with GPs (Chu & Ghahramani, 2005). The *movielens* dataset was sub-sampled from the original size of $\approx 1,000,000$ and zip codes were converted to latitudes and longitudes. The remaining datasets are available from the UCI Machine Learning Repository. In all experiments, data is normalized using training data only and the same normalization is applied to the test data.

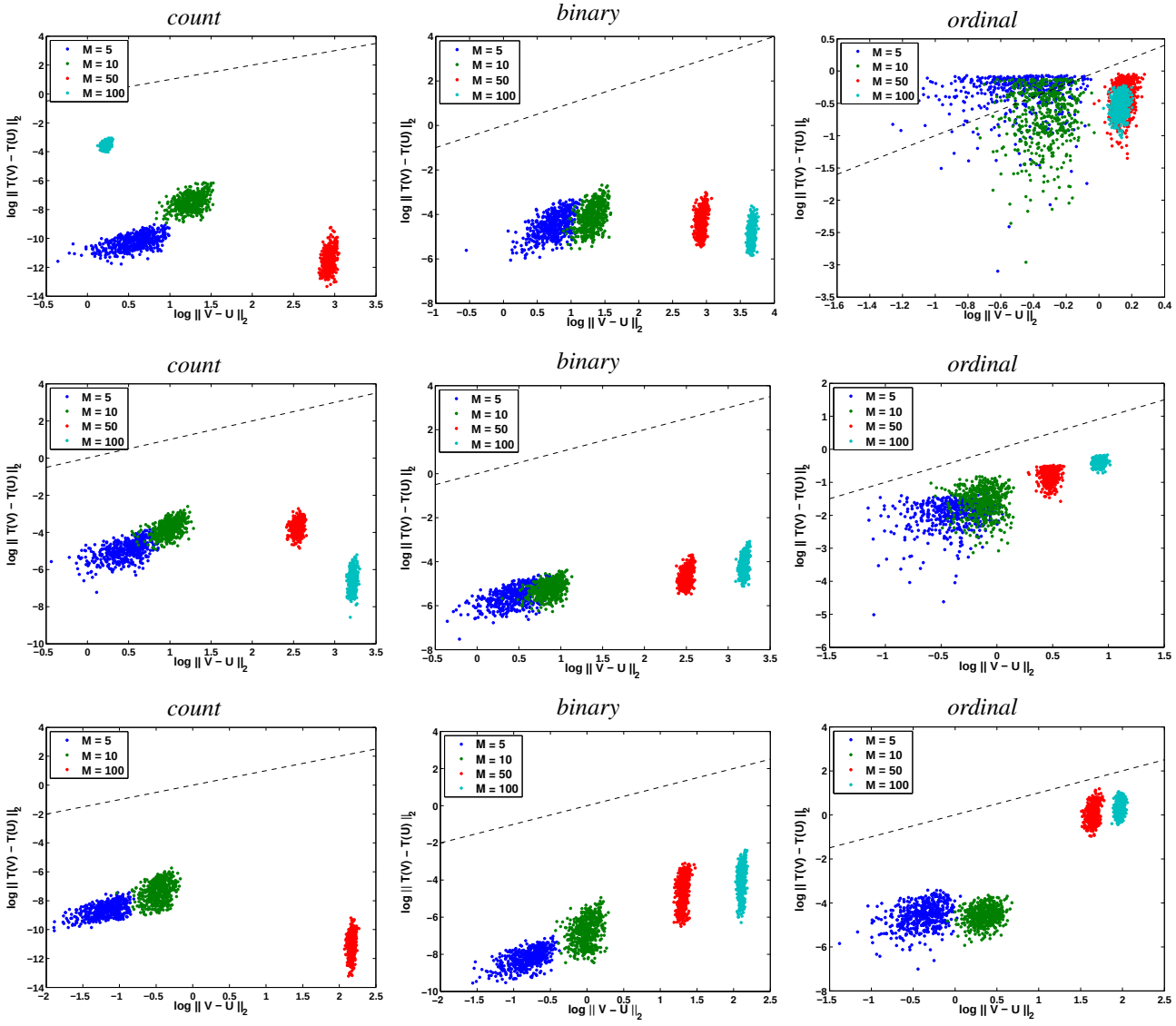

Figure 1. Results of contraction tests shown on log scale. The color coding refers to active set size. The dashed line represents the curve $\|T(U) - T(V)\|_2 = \|U - V\|_2$ in log space. The top row are the results from running with a zero-mean function and Gaussian RBF ($\ell = \frac{\sqrt{10}}{3}, \sigma^2 = 1$) kernel. The middle row are the results from running with a zero-mean function and Matern RBF ($\nu = \frac{1}{2}, \ell = \frac{\sqrt{10}}{3}, \sigma^2 = 1$) kernel. The bottom row are the results from running with a constant mean function (1) and 2nd degree polynomial ($c = 1$) kernel.

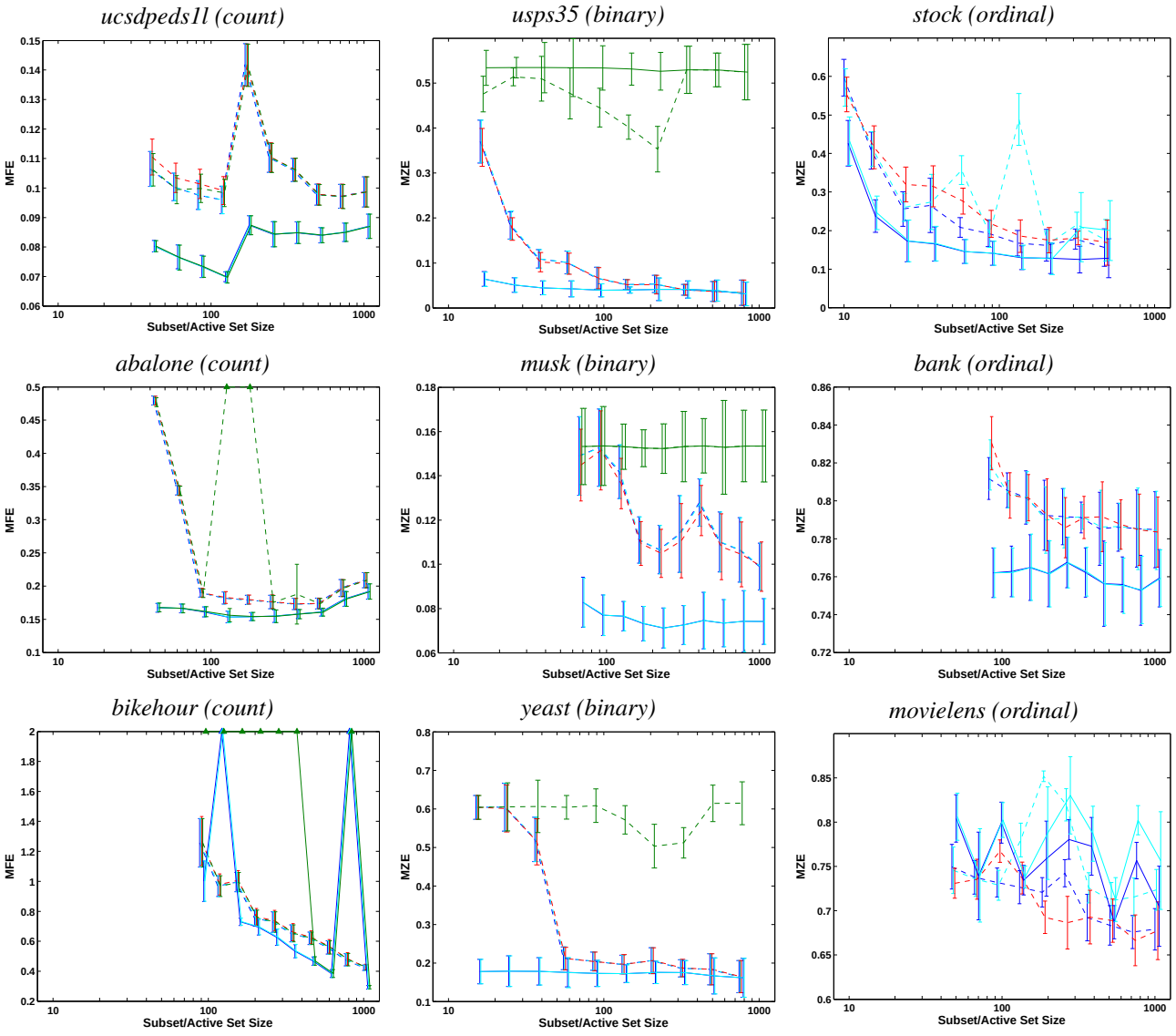

Figure 2. Learning curves with respect to subset/active set size. MFE is mean fractional error and MZE is mean zero-one error. Lower values represent better performance. Triangles on the edges of plots refer to data that exists outside the axes of the plot. Legend for plots: Laplace on SoD (- -), gradient ascent on SoD (- -), dual on SoD (- -), fixed point on SoD (- -), gradient ascent on full data (—), dual on full data (—), fixed point on full data (—).

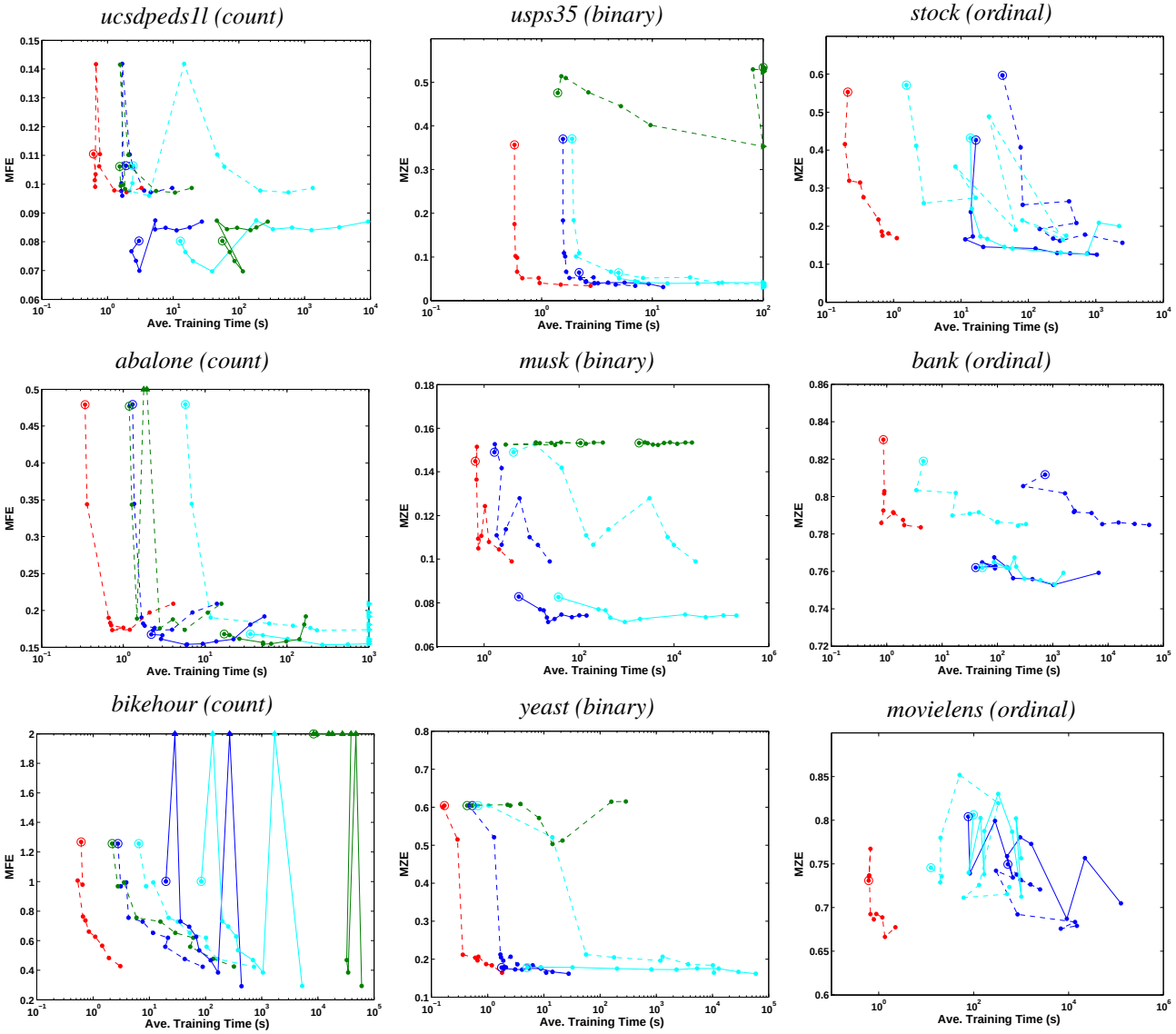

Figure 3. Training time / accuracy curves. Each dot represents a different subset/active set size. A circled dot represents the smallest subset/active set size for a method. Legend for plots: Laplace on SoD (- -), gradient ascent on SoD (- -), dual on SoD (- -), fixed point on SoD (- -), gradient ascent on full data (—), dual on full data (—), fixed point on full data (—).

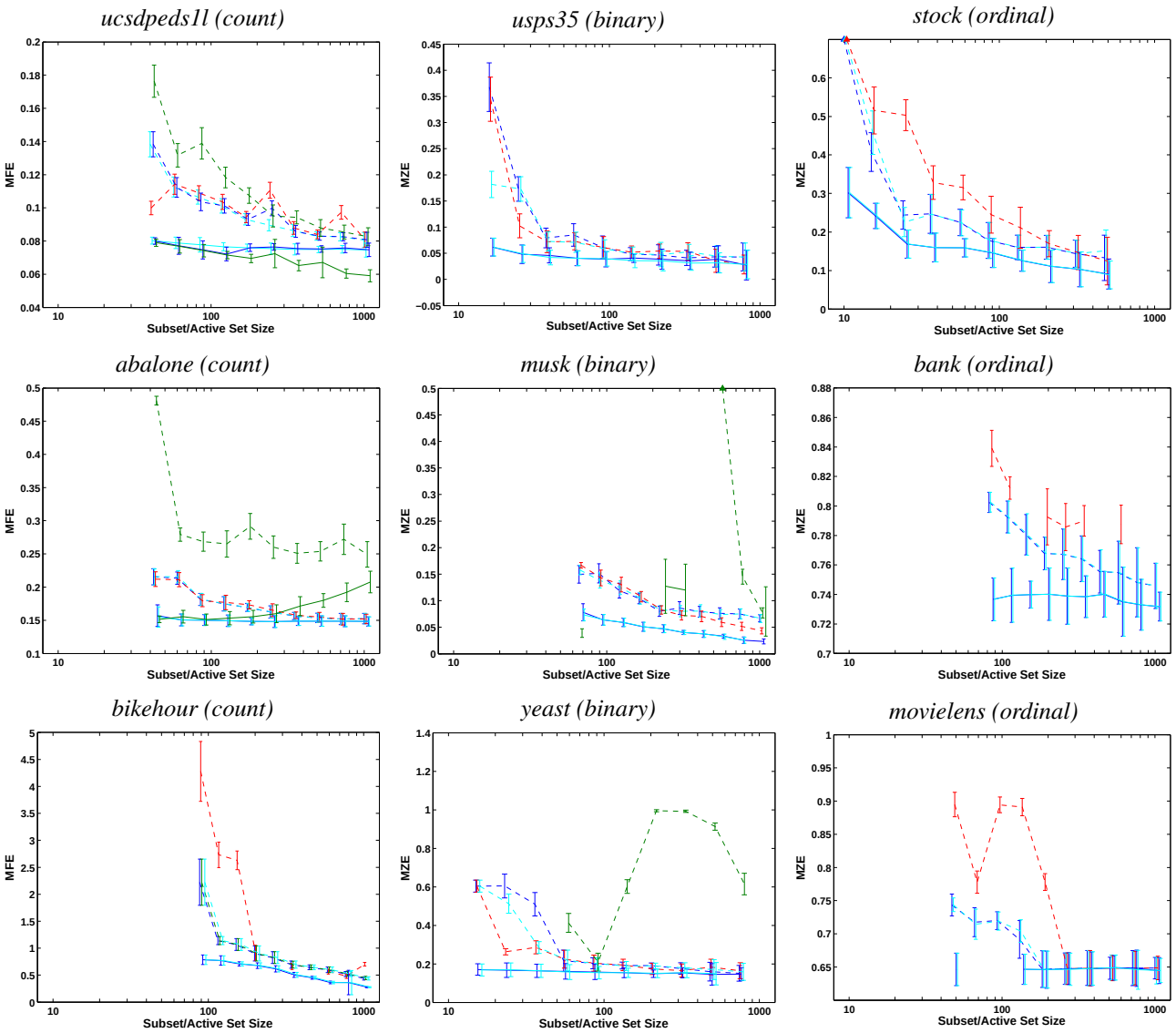

Figure 4. Results of hyperparameter optimization with respect to subset/active set size. A circled dot represents the smallest subset/active set size for a method. Legend for plots: Laplace on SoD (- -), gradient ascent on SoD (- -), dual on SoD (- -), fixed point on SoD (- -), gradient ascent on full data (—), dual on full data (—), fixed point on full data (—).

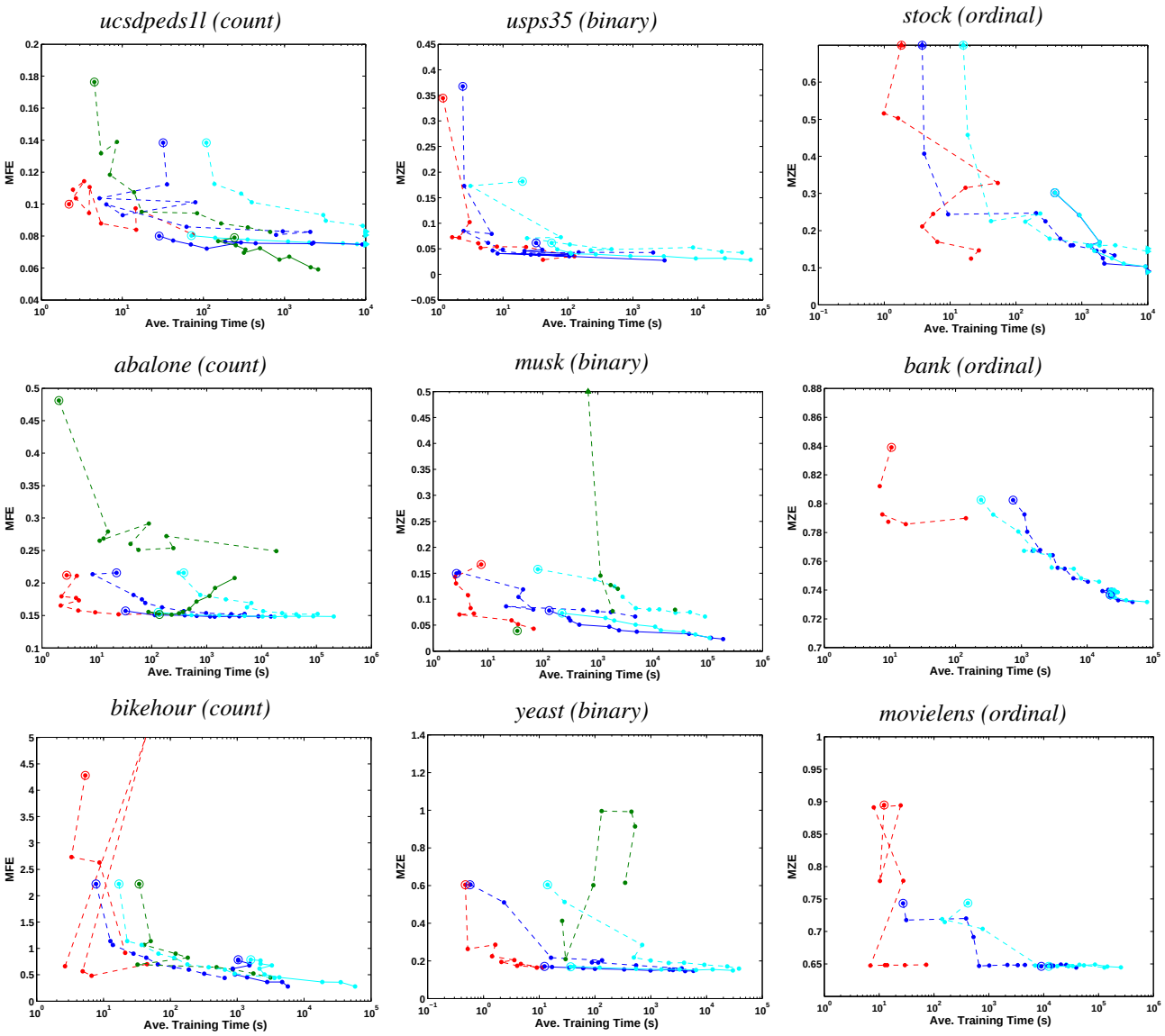

Figure 5. Results of hyperparameter optimization with respect to training time. Each dot represents a different subset/active set size. A circled dot represents the smallest subset/active set size for a method. Legend for plots: Laplace on SoD (- -), gradient ascent on SoD (- -), dual on SoD (- -), fixed point on SoD (- -), gradient ascent on full data (—), dual on full data (—), fixed point on full data (—).

References

Chan, A. B. and Vasconcelos, N. Counting People With Low-Level Features and Bayesian Regression. *Image Processing, IEEE Transactions on*, 21 (4):2160–2177, April 2012. doi: 10.1109/TIP.2011.2172800. URL http://dx.doi.org/10.1109/TIP.2011.2172800.

Chu, Wei and Ghahramani, Zoubin. Gaussian Processes for Ordinal Regression. *J. Mach. Learn. Res.*, 6: 1019–1041, December 2005. ISSN 1532-4435. URL http://portal.acm.org/citation.cfm?id=1088707.

Fanaee-T, Hadi and Gama, Joao. Event labeling combining ensemble detectors and background knowledge. *Progress in Artificial Intelligence*, 2(2-3):113–127, 2014.